THE UNIVERSITY *of* EDINBURGH

Biophysical Characterisation of the Hepatocyte Growth Factor - Glycosaminoglycan Interaction

Conny M Johansson

PhD Thesis

University of Edinburgh

2011

Abstract

Glycosaminoglycans (GAGs) such as heparin, heparan sulfate (HS), chondroitin sulfate (CS) and dermatan sulfate (DS) are sulfated polysaccharides that exist on animal cell surfaces and in the extracellular matrix. GAGs are important in providing structural and hydrating support and interaction points for proteins of varied functions, for example growth factors and homeostasis regulatory proteins. Hepatocyte Growth Factor (HGF) is a protein growth factor that regulates cell growth, survival, proliferation, chemotaxis, cell morphology, tissue regeneration and angiogenesis. It is involved in embryogenesis, wound healing and many cancers.

In this project, the interactions between the GAG binding N and NK -domains of HGF (HGF-N and HGF-NK) and different types of GAGs are characterised with biophysical techniques. GAG oligosaccharides were produced by enzymatic digestion and purified by preparative gel filtration and ion exchange chromatography. Different constructs of HGF were cloned from human cDNA, expressed with the *Pichia pastoris* expression system, purified to homogeneity and characterised by mass spectrometry and nuclear magnetic resonance (NMR).

The dissociation constants between the different HGF protein constructs, different heparin oligosaccharide lengths and the drug Fondaparinux were shown by isothermal calorimetry (ITC) to vary between 0.35 and 9.26 μM. It was found that the entropy contribution was favourable for short oligosaccharides and disfavourable for long oligosaccharides and that the enthalpy contribution was less important for shorter oligosaccharides than for longer oligosaccharides.

NMR titrations of CS, DS, heparin, Fondaparinux and sulfated maltose into ^{15}N labelled protein samples showed that all ligands bind to the same HGF-N binding site, but different binding modes exists. The binding site consists of three regions, with the α2-helix and L2 loops playing key roles (residues 70-84). All GAGs also utilise the N-terminal residues 32-42, whereas long heparin oligosaccharides can also utilise a binding region formed mainly by the β2-strand (residues 59-64, 66, 95, 96). The GAG binding mode changes if HGF-N has an N-terminal truncation and the β2-strand residues become more important, emphasising the role of the N-terminal residues in the HGF-GAG interaction. Spin-labelled fully sulfated heparin-derived hexasaccharide was used to determine its binding direction on the HGF-N surface. Affinity chromatography confirmed the importance of the N-terminal residues and that HGF binds to all investigated GAGs.

The oligomeric states of HGF-N and HGF-NK were investigated by AUC, gel filtration and ITC. The results suggest that the proteins oligomerise like beads on a string for long oligosaccharides. An HGF-N self-associating dimer with a slow on/off rate was characterised by affinity chromatography, gel filtration and NMR.

Acknowledgements

This thesis would not have been possible without the support and advice from a handful of people to whom I offer my gratitude.

I am very thankful to my supervisor, Dušan, for being so patient and caring. Your insight has been invaluable and your character has been a good balance to mine. I especially appreciate how much time you have devoted to my thesis.

Andy, I really appreciate the encouraging support you gave me throughout my PhD studies - you always cheered me up and provided good advice.

Bärbel, the help and advice you provided were invaluable. Your company, friendship, skill and drive always encouraged me.

Maria, I have honestly appreciated your company throughout my PhD, especially towards the end.

Carla, your energy, craziness and alternative vocabulary always cheered me up. I sincerely enjoyed getting to know you.

Christoph, you were always a helpful, organised and a true role model for a competent researcher.

Haris, I really like your calmness and relaxed attitude towards life, still being professional. You are not afraid of helping out and I really appreciate your assistance with all the practicalities around this thesis.

Mara, I always enjoyed our discussions and you cheered me up.

Felicia, I really appreciated our mini-project and our chats in the corridor. Your hard work and drive always inspired me.

Juraj, you are one of the most helpful and caring people that I have ever met and your NMR skills saved me many times.

Everyone that has contributed to the friendly and helpful spirit in the lab that has helped me produce this thesis: Patience; Nicky; Elisa; Isa; Marie; Dave; Dinesh; Paul; and anyone that I have forgotten to mention.

I would also like to thank all the support staff, John, Marika, the guys in stores and the postmen for taking care of all my online shopping. Janice for helping me with random stuff. Andy for showing me the MALDI-TOF.

Theresa, I have so much to be grateful for, but when it comes to my PhD, I deeply appreciate that you encourage me, care, understand and accept my feelings about it.

Världens största tack till Mamma, Pappa, Marie, Tobbe, mormor och alla andra hemma i Sverige för all hjälp med de mest absurda saker och för att ni alltid finns där även om vi ses så sällan och jag alltid är någon annanstans.

Unless stated in the text, the work described in this thesis is my own work and has not been submitted in whole or in part for a degree or other qualification at this or any other university.

Contents

Abbreviations

AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
AOX	alcohol oxidase
A_{xyz}	Absorbance at xyz nm
BMG	buffered minimal glycerol
BMGY	buffered minimal glycerol yeast extract peptone
BMM	buffered minimal methanol
BMMY	buffered minimal methanol yeast extract peptone
CCSD	combined chemical shift difference
CHO	cancer hamster ovary
CS	chondroitin sulfate
CV	column volume
dH_2O	distilled water
ddH_2O	double distilled water
DO	dissolved oxygen
DPn	oligosaccharide with degree of polymerisation n
DS	dermatan sulfate
E. coli	*Escherichia coli*
E_{XYZ}	extinction coefficient at XYZ nm
ECD	electron capture dissociation
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
EPR	electron paramagnetic resonance
EtOH	ethanol
FC	flow cell
FGF	fibroblast growth factor
FPLC	fast protein liquid chromatography
FTICR	fourier transform ion cyclotron resonance
GAG	glycosaminoglycan
GlcA	glucuronic acid
GlcNAc	N-acetylated glucuronic acid
GlcNS	N-sulfated glucuronic acid

GMSA	gel mobility shift assay
GPI	glycosylphosphatidylinositol
HAI	hepatocyte growth factor activator inhibitor
HGF	hepatocyte growth factor/scatter factor
HGF-N	N-terminal domain of HGF
HGF-N-Δ6	N-terminal domain of HGF lacking the first 6 residues
HGF-K	kringle 1 domain of HGF
HGF-NK	N-terminal domain and kringle 1 domain of HGF
HGF-NK-C	HGF-NK cloned in Edinburgh by Conny Johansson
HGF-NK-G	HGF-NK cloned in Cambridge by the Gherardi group
HPLC	high performance liquid chromatography
HS	heparan sulfate
HSPG	heparan sulfate proteoglycan
HSQC	heteronuclear single quantum correlation
IdoA	iduronic acid
ITC	isothermal titration calorimetry
kDa	kilodalton
LB	lysogeny broth
M	mass
MM	minimal medium
MW	molecular weight
MeOH	methanol
MWCO	molecular weight cut off
m/z	mass/charge ratio
NA	N-acetylated
NS	N-sulfated
NDST	N-deacetylase/N-sulfotransferase
NK1	N-terminal domain and kringle 1 domain of HGF
NMR	nuclear magnetic resonance
OD600	optical density at 600 nm
PAPS	3-phosphoadenyl 5-phosphosulfate
P. Pastoris	*Pichia pastoris*
PBS	phosphate buffered saline

PDGF	platelet-derived growth factor
pI	isoelectric point
PG	proteoglycan
PMSF	phenylmethyl sulfonyl fluoride
PTM	posttranslational modification
RD	relexation delay
$(RG)_{11}$	RGRGRGRGRGRGRGRGRGRG peptide
rpm	revolutions per minute
SAR	structure activity relationship
SAX	strong anion exchange
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel
SOS	sucrose octasulfate
SPH	serine protease homology domain
SPR	surface plasmon resonance
SV	sedimentation velocity
TCEP	tris[2-carboxyethyl]phosphine hydrochloride
TEMPO	2,2,6,6-tetramethylpiperidine-1-oxy
TEMPO-DP6	fully sulfated heparin DP6 conjugated to TEMPO by reductive amination
TFA	trifluoroacetic acid
UV/Vis	ultraviolet/visible
VEGF	vascular endothelial growth factor
YNB	yeast nitrogen base
YPD	yeast extract peptone dextrose
YPDS	yeast extract peptone dextrose sorbitol
YPDSZ	yeast extract peptone dextrose sorbitol zeocin
Z	charge

1 Introduction

1.1 Background

Hepatocyte Growth Factor (HGF) is a multifunctional growth factor that controls many physiological processes such as cell growth and morphology [1]. There is extensive evidence of its involvement in cancer and wound healing. HGF exerts its biological functions by binding the cell-surface receptor Met and a glycosaminoglycan (GAG) at the same time [2]. Because HGF is involved in many physiological processes and its dysregulation can cause disease, it is an important research target.

GAGs are large linear and complex carbohydrates composed of a basic repeating disaccharide unit, which is extensively modified. They either exist as free carbohydrates or are covalently linked to proteins. Some types of GAGs bind water and work as a structural scaffold between cells, for example in cartilage and joints [3] and for a long time it was thought that the only function of GAGs was to provide structural support. However, during the last decades it has been realised that GAGs are also involved in complex signalling events such as the HGF signalling cascade. GAGs can bind to signalling proteins such as growth factors and modulate how they interact with their receptors. Protein – GAG interaction can therefore be targeted for drug development [4]. Heparin anticoagulant therapy is one treatment that targets a protein-GAG interaction that has been used for 7 decades. However, a new generation of innovative anti-cancer drugs are currently being developed by targeting the interaction between GAGs and growth factors [5-8].

In order to target protein-GAG interactions for drug development and design effective drug molecules by structure-activity-relationships (SAR), an in-depth understanding of the interaction on a molecular level is needed. Many different aspect of HGF have been researched in detail, for example its cancer biology [9]. A number of biophysical studies of HGF are available, for example the affinity of the interaction has been determined by SPR [10-13], the structure of some domains of

the protein has been determined by x-ray crystallography [14-16] and NMR [17] and the GAG binding site has been mapped.

However, there are gaps in the understanding of this important protein-GAG interaction. For example, a comparison of the binding modes of different types of GAGs is lacking and the oligomerisation mode of some HGF protein constructs is still unclear. In addition, most of the current understanding of the HGF-GAG interaction is derived from a crystal structure in which the protein contains a cloning artefact. The effect of this cloning artefact and how it affects the GAG binding mode has previously not been investigated. Several solution based biophysical techniques are used in this thesis to further characterise the HGF-GAG interaction in order to aid future HGF basic research and drug development.

1.2 Hepatocyte Growth Factor

Growth factors are extracellular signalling proteins that bind cell surface receptors to stimulate cell growth, survival or proliferation [18]. Hepatocyte growth factor (HGF) is extraordinary among the growth factors in the way that it has been shown to regulate all these activities, plus chemotaxis, cell morphology, tissue regeneration and angiogenesis [1]. HGF has also been called hepatopoietin and scatter factor, although HGF is the current standard terminology. HGF was first cited in 1980 and since it was shown that HGF is the ligand for the receptor tyrosine kinase Met, it has been extensively studied for its role in cancer and tissue regeneration.

1.2.1 The Structure of Hepatocyte Growth Factor

HGF, ExPASy entry p14210, consists of one N-terminal PAN domain (abbreviated N), followed by four kringle domains (abbreviated K1-K4) and lastly a serine protease homology domain (abbreviated SPH). See Figure 1.1 for a schematic structure of HGF and Table 1.1 for an extract from ExPASy. The serine protease homology domain is not catalytically active; two of the inactivating changes affecting the catalytic triad are His534Gln and Ser673Tyr. Abbreviations are often

used for splice variants and fragments of HGF, for example NK2 consists of N, K1 and K2. In this project, HGF-N denotes the N domain of HGF, HGF-NK, the N and the K1 domains, HGF-NK-C denotes the NK fragment that was cloned in this project and HGF-NK-G, the NK fragment that was originally cloned in Cambridge by the Gherardi research group. NK4 forms the α-chain, whereas the SPH domain forms the β-chain after proteolytic cleavage by a protease. Residues 1-31 make up the secretory signal that directs the protein for secretion via the endoplasmatic reticulum and Golgi apparatus. The secretory signal is cleaved off and not present in the native extracellular protein. NK1 has been detected *in vivo* as a splice variant of HGF [19]. In crystal structures, NK1 forms dimers [14-16], [20], but it is monomeric in solution [21].

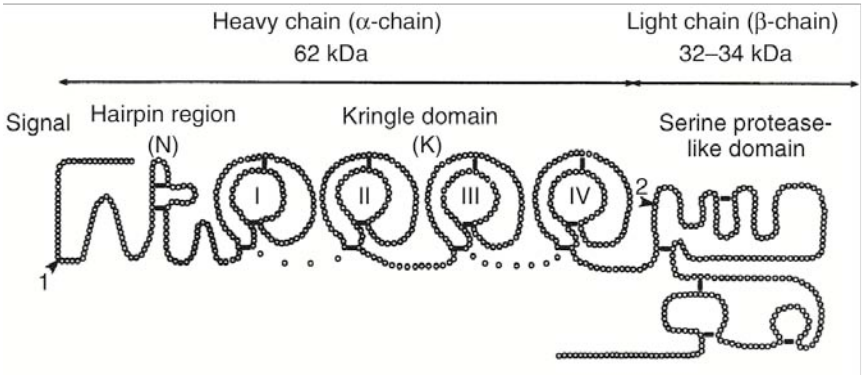


Figure 1.1 Schematic structure of HGF. The first arrow is the signal peptide cleavage site. The second arrow denotes the activation cleavage site. Taken from [22].

Table 1.1 Extract of HGF (P14210) from the ExPASy database.

Function	From – To	Length	Domain type
SIGNAL	1 - 31	31	Signal peptide.
CHAIN	32 - 494	463	Alpha chain.
CHAIN	495 - 728	234	Beta chain.
DOMAIN	37 - 123	87	PAN.
DOMAIN	128 - 206	79	Kringle 1.
DOMAIN	211 - 288	78	Kringle 2.
DOMAIN	305 - 383	79	Kringle 3.
DOMAIN	391 - 469	79	Kringle 4.
DOMAIN	495 - 721	227	Peptidase S1.

The structure of HGF has been extensively studied, there is one NMR structure of the N domain [17], several crystal structures of free NK1 and NK1 complexed with heparin [14-16] as well as crystal stuctures of free HGF-β and HGF-β complexed with the Met Sema domain [23], [24]. There is also a crystal structure of NK2 [25]. This crystal structure is shown in Figure 1.2, overlaid with the crystal structure of NK1 complexed with heparin. The secondary structure of the N domain is shown in Figure 1.3.

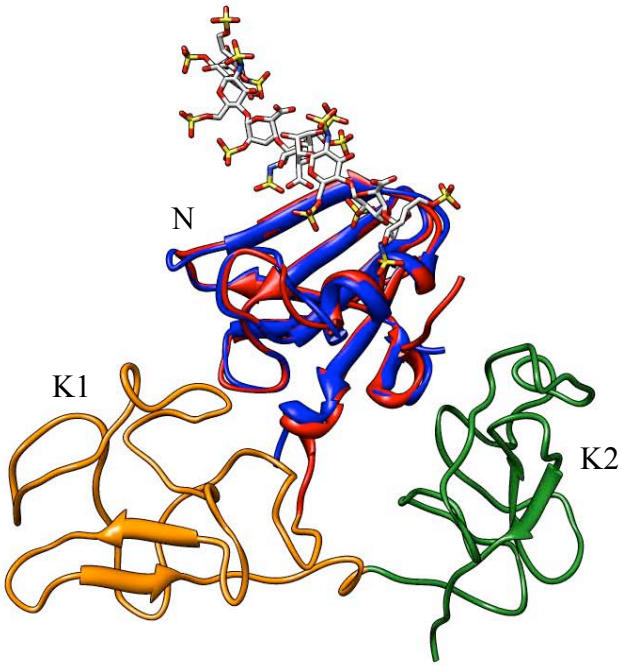


Figure 1.2 Overlay of the 3HN4 NK2 PDB file and the 1GMO HGF-NK-G – heparin complex PDB file. Blue: 1GMO N-domain, Red: 3HN4 N-domain, orange: 3HN4 K1-domain, green: 3HN4 K2-domain.

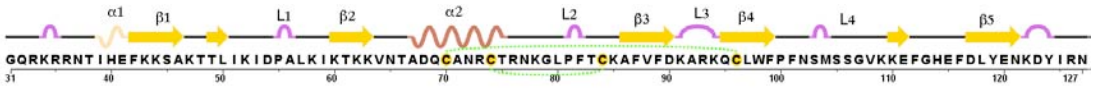


Figure 1.3 The secondary structure of the N domain of HGF as determined in the 2HGF PDB file.

1.2.1 Molecular Physiology of Growth Factors

Many growth factors bind receptor tyrosine kinases (RTKs) and cause the receptors to dimerise or oligomerise [26]. As illustrated in Figure 1.4, the intracellular tyrosine kinase domains of the receptors then cross-phosphorylate each other and the phosphorylations are recognised by adaptor proteins that trigger an intracellular signalling cascade. The role of proteoglycans and GAGs in this model is to attract the growth factors to the cell surface and present them to the RTKs.

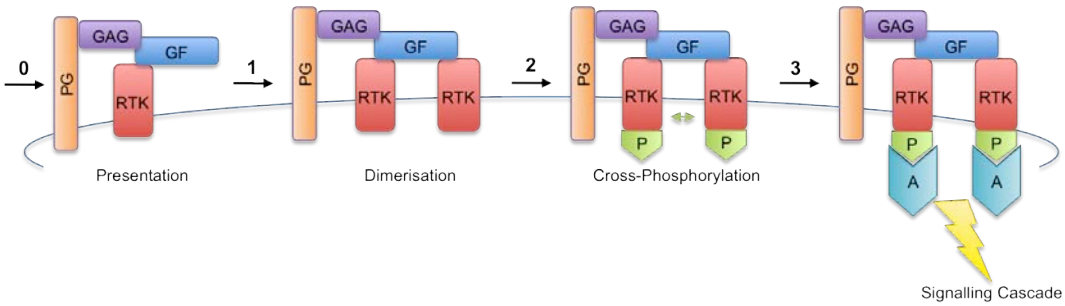


Figure 1.4 Typical activation event of an RTK. 0) Presentation of GF to RTK by PG/GAG. 1) Dimerisation of RTK by GF and GAG. 2) RTK cross-phosphorylation. 3) Binding of adaptor proteins and initiation of signalling cascade. Abbreviations: RTK: Receptor Tyrosine Kinase, PG: Proteoglycan, GAG: Glycosaminoglycan, GF: Growth Factor, P: Phosphorylation, A: Adaptor Protein.

Different growth factors employ different types of oligomerisation strategies. For example, PDGF is a protein homo-dimer and can thus bind two receptors [27]. FGF1, its receptor FGFR2c and heparin oligosaccharides form a 2:2:1 complex with the heparin acting like glue, making the FGF1 bind in a *trans* manner [28-30]. On the other hand, it has been shown by NMR that a biologically active hexasaccharide makes FGF1 dimerise in a *cis* or *trans* manner and it is proposed that FGF *trans* dimerisation is not a prerequisite for receptor activation [31]. There are several isoforms of EGF receptors and the composition of the heterodimeric receptors partly regulates the activation of the signalling complex [32-34]. For HGF, the stoichiometry of the active signalling complex still remains unclear [23-25], [2].

1.2.2 The Hepatocyte Growth Factor – Met Signalling Complex

It has been postulated that HGF has one high affinity Met binding site in the α-chain and one lower affinity site in the β-chain, making HGF bivalent for Met and thus being able to dimerise the Met receptor [24], [23], [35]. This suggest that the active signalling complex has a 2 Met : 1 HGF stoichiometry, although the stoichiometry has not been completely clarified and a crystal structure of full-length HGF, Met and heparin is needed to ultimately answer the stoichiometry question. Analytical

ultracentrifugation and light-scattering experiments suggest that active Met and HGF form 1:1 complexes that can be stabilized with heparin to form a 1:1:1 complex [36]. No 2:1 complexes were detected in these experiments, but the absence of this stoichiometry could be due to the fact that at the 1:1 molar ratio of ligand and receptor used in these experiments, only high-affinity HGF α-chain:Met interactions are taking place, and low-affinity HGF β-chain:Met interactions are not taking place and therefore, 2:1 complexes cannot form.

One could also imagine a 2:2 complex of Met:HGF where HGF binds in a dimeric fashion [2]. This model is partly supported by the fact that HGF-NK crystallises as a dimer, on the other hand it is monomeric on solution. The 2:1 and 2:2 complexes are illustrated in Figure 1.5

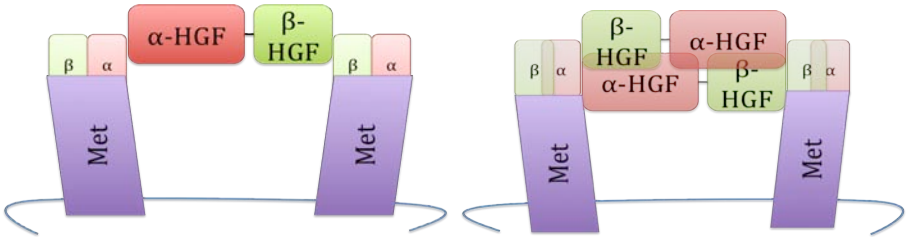


Figure 1.5 Two different dimerisation models for the HGF-Met complex. A 2:1 model on the left and a 2:2 model on the right.

It has been shown that the presence of a HS like GAG potentiates the activity of HGF although it is not necessary [2], [37-39], [10]. DS and CS like GAGs have also been shown to potentiate HGF activity [40], [37], [41]. On the other hand, NK1 only binds Met in the presence of heparin. It is mainly the N domain that interacts with GAGs although some GAG binding is also seen for K1 and K2 [42].

Full-length HGF can bind Met with similar affinities regardless if it is activated or not, although only activated HGF can activate Met [43]. Both the N and SPH -

domains have been shown to interact with Met [42], with the N domain binding tighter than the SPH domain [23], although the SPH domain has been shown to be vital for Met signalling [24]. By mutating residues in the SPH domain activation pocket, which are necessary for HGF activation, it is possible to convert the β-chain of HGF into a Met antagonist [44].

Interestingly, HGF-NK1 is a Met agonist, crystallises as a head-to-tail dimer, but by mutating residues in the HGF-NK1 dimer interface, it is possible to convert it into a Met antagonist [45]. This is likely due to the fact that wild-type HGF-NK1 is able to dimerise the Met receptor as shown in Figure 1.5. HGF-NK2 crystallises in a globular monomeric fashion due to interactions between the N and K2 domains and works as a Met antagonist, likely through binding to Met and inhibiting Met dimerisation [25]. Mutations that were to open up the HGF-NK2 closed conformation by disrupting the N/K2 interface convert HGF-NK2 from a Met antagonist to an agonist. Remarkably, this mutated HGF-NK2 agonist can be converted back to an antagonist by a mutation that disrupts the NK1/NK1 dimer interface.

What role a HS-like GAG would have in the active signalling complex is still unclear. It is however likely that a PG linked GAG contributes to presenting HGF to the Met receptor. It is possible that a GAG is also present in the active signalling complex, as suggested by the finding that HGF-NK1 needs heparin to bind Met and that Met has been shown to bind heparin [39], [37]. It has been suggested that the primary role of heparin in HGF-NK1 binding to Met is to neutralize the positive charged residues on the N-domain surface. Removal or charge reversal of heparin binding residues within the N-domain makes HGF-NK1 bind to Met independently of heparin [25]. Decorin, a secreted matrix proteoglycan belonging to the small leucine-rich proteoglycan family has been shown to bind directly to Met with a K_D of 1.5 nM [46]. Decorin bound to Met is displaced by HGF as it binds to Met with higher affinity. It has also been suggested that HGF binds Met as a monomer, but that heparin/HS is needed as a co-receptor to induce a conformational change [37].

Although HS and heparin have higher affinity for HGF than dermatan sulfate, they have been shown to be equally potent co-activators of Met [37].

Some cell-based assays have shown that GAGs are required to activate Met, others have shown that HGF alone is also active. Clearly, results seem to vary with cell type and experimental setup, but there is general agreement that GAGs are important as co-activators [2].

1.2.3 Activation of Hepatocyte Growth Factor by Hepatocyte Growth Factor Activator

For HGF to be able to activate Met, it first has to be activated by proteolytic cleavage between residues Arg494 and Val495 [47]. The activation site can be knocked out by a R494E point mutation. Some of the activating proteases can also process a secondary site between Arg424 and His425 in the K4 domain. After the proteolytic cleavage, the α and β -chains are attached by a disulfide bond. It has been shown that many proteases such as coagulation factor XIa, XIIa, kallikrein, urokinase, and tissue-type plasminogen activator can perform the cleavage [47], but HGFA [48] and matriptase [49], [50] have been identified as the most potent and physiologically relevant activators. Both HGFA [51] and matriptase [50] are serine proteases that have been implicated in cancer. Matriptase is a transmembrane protein that also degrades extracellular matrix proteins such as gelatin, fibronectin and laminin [52]. Matriptase is thought to be a less efficient HGF activator than HGFA, but as it is often overexpressed in cancer, it is still thought to be physiologically relevant.

The proteolytic cleavage leads to rearrangements in the activation domain of β-HGF and the reorientation of the β-chain relative to the α-chain. The activation domain consists of three surface-exposed loops and the newly formed N-terminus. It is thought that insertion of the new N-terminus into the activation pocket is related to the activation mechanism of other trypsin-like serine proteases [44]. This rearrangement results in the formation of what is the active site in other trypsin-like serine proteases, but for HGF it interestingly forms the Met binding site. It has been

shown that single-chain pro-HGF adopts a compact conformation, whereas activated HGF is elongated [53].

1.2.4 The Structure and Molecular Physiology of Hepatocyte Growth Factor Activator

HGFA is secreted mainly by the liver and has been shown to be an important regulator of wound healing [22]. The crystal structure of HGFA has been solved, both in its free form and complexed with a Hepatocyte Growth Factor Inhibitor peptide [54]. In its free state, HGFA adopts an active site conformation that differs from many other trypsin-like proteases and the catalytic triad cannot function. Complexed with a HAI-1B peptide, HGFA undergoes a major conformational change and the resulting state closely mirrors the active site of other serine proteases, but the complex is inactive [54], [55].

HGFA is secreted as a 96 kDa zymogen with a domain structure similar to coagulation factor XII, consisting of six domains: an N-terminal fibronectin type II domain; an epidermal growth factor like domain; a fibronectin type I domain; another epidermal growth factor like domain; a kringle domain; and a trypsin homology serine protease domain, all together forming the pro-peptide that is removed by plasma kallikrein cleavage at the Arg372-Val373 bond, see Table 1.2 and Figure 1.6. The most C-terminal domain is the trypsin homology serine protease domain that can be activated by cleavage of the Arg407-Ile408 bond by thrombin or kallikrein 5 in the presence of negatively charged molecules such as dextran sulfate and chondroitin sulfate. Alternatively, HGFA can be activated by kallikrein 4 that does not need negatively charged molecules to be present. Consequently, the active short form of HGFA is a 34 kDa protein containing a C-terminal 35-residue peptide linked by a disulfide to the protease domain [56], [57]. Thrombin can also activate the long 96 kDa form of HGFA, before it has been cleaved by plasma kallikrein (see Figure 1.7 for the different activation pathways of HGFA). Kallikrein 5 is a more efficient pro-HGFA activator than kallikrein 4 and has been linked to cancer [58].

Table 1.2 Extract of HGFA, Q04756 from the ExPASy database.

Function	From - To	Length	Domain type
Signal peptide	1 – 35	35	Secretion signal
Propeptide	36 – 372	337	Removed in mature form
Chain	373 – 407	35	HGFA short chain
Chain	408 – 655	248	HGFA long chain
Domain	103 – 150	48	Fibronectin type-II
Domain	160 – 198	39	EGF-like 1
Domain	200 – 240	41	Fibronectin type-I
Domain	241 – 279	39	EGF-like 2
Domain	286 – 367	82	Kringle
Domain	408 – 646	239	Peptidase S1
Active site	447	1	Catalytic triad
Active site	497	1	Catalytic triad
Active site	598	1	Catalytic triad

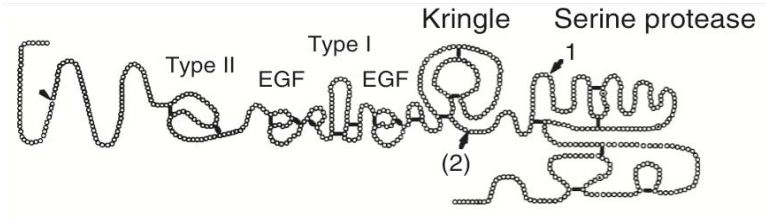


Figure 1.6 Overall structure of HGFA and the proteolytic cleavage sites. Cleavage site 1: Plasma kallikrein. Cleavage site 2: Thrombin. Taken from [22].

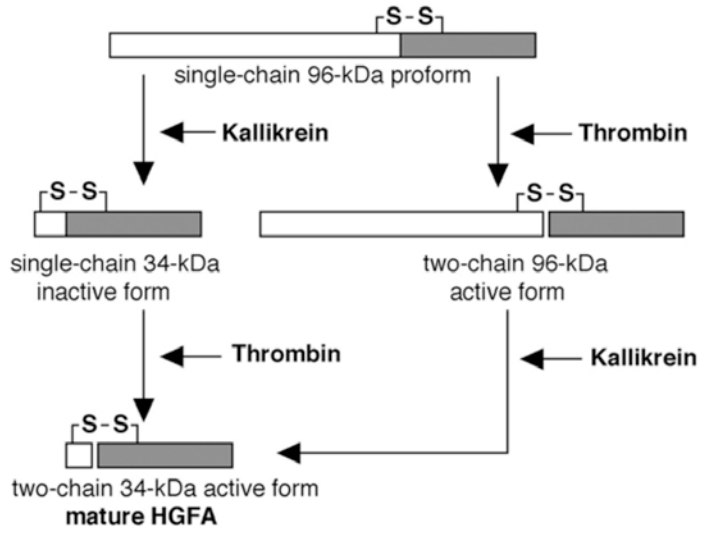


Figure 1.7 The activation pathway of HGFA. Taken from [52].

Kunitz-type serine protease inhibitors called hepatocyte growth factor inhibitors (HAIs) regulate HGFA activity. Despite the name, HAIs can also inhibit other proteases such as matriptase, plasma kallikrein and kallikrein 4 and 5 [59]. The HAIs are expressed as transmembrane proteins. The primary HAI-1 translation product is a 66 kDa protein, but the ectodomain can be shedded through proteolytic cleavage by metalloproteases at different cleavage sites [60]. This release of the inhibitors could decide the function of HAI-1 and HAI-2 [61]. It is the 50 amino acids long Kunitz domain of the HAIs that bind to HGFA and cause inhibition [54].

The fact that HGFA is activated by thrombin, which is activated by tissue injury and the blood coagulation cascade, strongly links the HGFA/HGF/Met signalling pathway to tissue injury [22], [62]. HGFA is found in its active form in serum, but in its inactive form in plasma. Interestingly, it has been shown that unactivated HGFA has weak affinity for heparin but acquires stronger heparin affinity after activation [62]. This feature likely ensures immobilisation of the active enzyme at the site of injury. The HGFA short chain contains several basic residues and has a high pI, it is therefore likely that the heparin binding properties reside in this part of the protein and that the proteolytic cleavage mobilises the short chain, making it able to interact

with GAGs. It is noteworthy that HGFA, HGF and Met all exhibit some degree of heparin affinity and this points to the central role of GAGs in the activation and regulation of the HGFA/HGF/Met signalling pathway as shown in Figure 1.8.

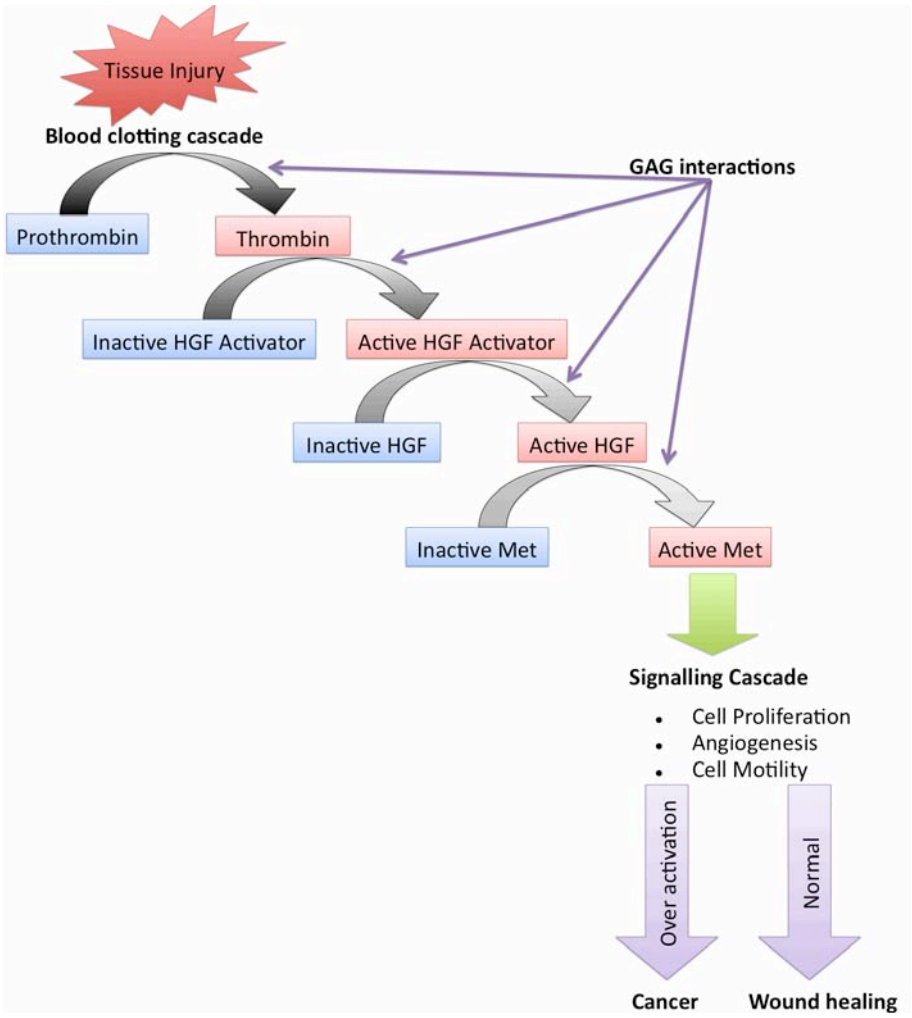


Figure 1.8 The central role of GAGs in the HGF signalling cascade.

1.2.5 The Hepatocyte Growth Factor - Glycosaminoglycan Interaction

The GAG binding site of HGF was first identified by mutational studies [11], confirmed by NMR [17], [21] and studied further by crystallising HGF-NK-G with a heparin-derived 14-mer oligosaccharide [14]. The heparin-binding site consists of a patch of positively charged arginines and lysines that are thought to form ionic intermolecular bonds with the negatively charged sulfate groups on the GAG, but

hydrogen bonds are also important. The most important protein residues in the interaction are the very N-terminal basic residues that cannot be seen in the structures [63], the residues of the β2-strand (Lys58, Lys60, Thr61, Lys62 and Lys63), the residues of the α2-helix (Arg73, Arg76 and Asn77) and the residues of the L2-loop (Lys78, Gly79 and Leu80) [14], [21]. Arg73 plays a central role in the interaction.

Although only one arginine - sulfate interaction is annotated as an ionic interaction in the published HGF-NK-G - heparin structure [14] (Table 1.3), most of the arginine/lysine - sulfate intermolecular bonds are likely to be of an ionic positive - negative charge character rather than a hydrogen bond donor - acceptor character. Indeed, most of the arginine/lysine - sulfate interactions in the thrombin-antithrombin-heparin complex are classified as ionic bonds or salt bridges [64]. A possible explanation to the differences in nomenclature is that many of the HGF - heparin hydrogen bond donor - acceptor distances are 2.5-3.3 nm long and on this scale, both hydrogen and ionic bonding can take place [65]. Both the distance and the angle of the bond donor and acceptor are important in determining the strength of a hydrogen bond and the optimal distance and angle depends on the specific system. Some interactions between the protein backbone amides and heparin sulfates occur in the HGF-NK-G - heparin co-crystal and these intermolecular bonds are of a hydrogen bond character as the amides are polar but not charged. Figure 1.9 shows a schematic figure of the protein - heparin interactions with bond lengths.

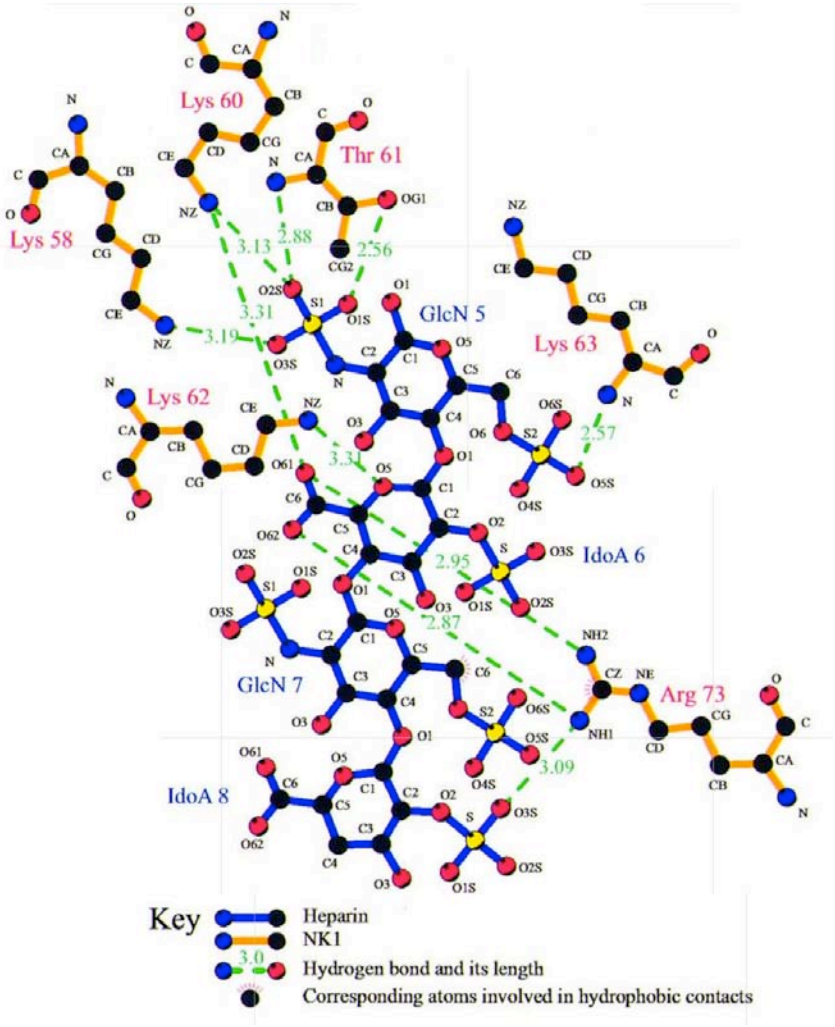


Figure 1.9 HGF-NK-G – heparin interactions. Taken from [14].

The intermolecular bonds between HGF-NK-G and heparin seen in the crystal structure are summarised in Table 1.3, using the terminology in the publication.

Table 1.3 Intermolecular interactions between HGF-NK-G and heparin as reported in the crystal structure [14].

GAG interaction point	**Type of bond**
GlcN 5 6-O sulfate	
Lys63 (backbone nitrogen)	Hydrogen bond
GlcN 5 N sulfate	
Thr61 (side chain oxygen)	Hydrogen bond
Thr61 (side chain oxygen + backbone nitrogen)	Hydrogen bond
Lys60 (side chain nitrogen)	Hydrogen bond
IdoA 6 COOH Oxygen 1	
Lys60 (side chain nitrogen)	Hydrogen bond
Arg73 (side chain nitrogen)	Hydrogen bond
IdoA 6 COOH Oxygen 2	
Arg73 (side chain nitrogen)	Hydrogen bond
IdoA 6 sugar ring O	
Lys62 (side chain nitrogen)	Hydrogen bond
IdoA 8 2-O sulfate	
Arg73 (side chain nitrogen)	Hydrogen bond
GlcN 7 carbon 6	
Arg73 (side chain carbon)	Hydrophobic
GlcN 7 6-O sulfate	
Arg73 (side chain nitrogen)	Ionic interaction
Gly79 (backbone nitrogen)	Polar

It has been suggested that 6-O-sulfates in GlcN units are the most important GAG interaction points for HGF, with IdoA 2-O-sulfates and N-sulfates playing minor roles [66]. As can be seen in Figure 1.9 and Table 1.3, there is only one intermolecular bond between the IdoA2S and HGF. The GlcN sulfates and in particular the 6-O-sulfates are involved in several intermolecular bonds, suggesting an important contribution and supporting the hypothesis that GlcN 6-O-sulfates are important interaction points. It has also been suggested that HGF has a general preference for IdoA containing GAGs [67]. This is likely due to the flexibility of IdoA allowing for better alignment of the protein-GAG intermolecular bonds and a similar preference is seen for many other GAG binding proteins.

A recent study using specifically desulfated heparin tetrasaccharides showed that species containing only two sulfates (N, 2O or 6O –sulfates) bind to HGF with equivalent affinities, as did a disulfated DS tetrasaccharide and it was suggested that "Affinity is dictated by the need for a minimum of an IdoA residue and two sulfates, which can be positioned in a variety of ways, located within a dp3/dp4 segment of either a HS/heparin or DS backbone" [68].

Although the HGF-GAG interaction clearly exhibits some specificity, it has been shown that no specific sulfates or oligosaccharide motif in a panel of DS and HS like GAGs is needed for HGF interaction and activity [67]. As often seen in protein - GAG interactions, different species of similar sulfate density have comparable properties and affinities and Met/HGF activation potencies increase with increasing sulfate densities [69]. This is exemplified by the fact that chemically sulfated maltose oligosaccharides have similar properties as heparin in cell based assays [38].

A tetrasaccharide has been found to be the shortest heparin fragment that binds to HGF-NK1 and activates Met, but with DS a hexasaccharide is required [70]. In the crystal structure, a tetrasaccharide is the minimal fragment that fully occupies the whole binding site. The K_D of the interaction does not change with the length of the

GAG, but GAGs longer than 10 sugar units are more active in cell based assays [10]. Although it seems like HS and heparin have higher affinity for HGF than DS, they have been shown to be equally potent activators for Met [37].

The K_D for the interaction between GAGs, HGF and its NKx splice variants have been determined mostly by biosensor analysis to be in a low nM range [10-13]. The GAG K_Ds for full-length HGF, HGF-N, HGF-NK1 and HGF-NK2 are all similar. Surprisingly few studies have employed alternative methods to assess the K_D of the HGF-GAG interaction. One study used FITC labelled heparin and fluorescence spectroscopy to establish that the K_D for the HGF-N - heparin interaction is 1-2 μM [21]. This significantly weaker K_D might partly be explained by experimental differences. The fluorescence spectroscopy experiment was performed in solution and the heparin was not immobilised. The incorporation of a FITC label, different sources and length of GAGs and different buffers might further explain the differences in experimental K_Ds. The binding affinity between NK1 and a fully sulfated heparin-derived tetrasaccharide (DP4C) in PBS has been determined by ITC to be in the micromolar range (Bärbel Blaum, University of Edinburgh, unpublished results), adding to the uncertainty of the actual strength of the interaction.

Measured NMR ^{15}N spin relaxation parameters of HGF-N indicate that parts of the protein undergo conformational exchange on a μS to mS timescale [21]. This NMR study also showed several new crosspeaks from the N-terminal residues and Asn77-Leu80 loop appearing in ^{1}H-^{15}N HSQC spectra due to the protein becoming more rigid upon GAG binding. Chemical shift changes were also seen for residues lying outside the central binding site, indicative of a possible structural rearrangement upon ligand binding. Changes in protein conformation upon GAG binding have been reported for other heparin binding protein such as TSG-6 [71]. Changes in ^{15}N spin relaxation parameters upon GAG binding have been reported for FGF [72]. The seven N-terminal residues are not observable by NMR spectroscopy due to conformational exchange [21]. This is also likely why the N-terminal residues cannot be seen in crystal structures.

1.2.6 HGF in Cancer and as an Anti-Cancer Target

HGF was first detected as a protein in rat serum that could increase hepatocyte DNA synthesis and was subsequently cloned and expressed [73], [74]. It was later shown that HGF binds the tyrosine kinase receptor Met [75], which was also characterised in the 1980s as a proto-oncogene [76]. Since then, the chemistry and biology of HGF and its relation to cancer have been extensively studied [77].

The HGFA-HGF-Met pathway regulates cell functions such as cell growth, apoptosis, motility and invasion, all hallmarks of cancer [78]. Dysregulation of this important pathway can occur at different levels: the cell can start over-expressing its own HGF, becoming self-sufficient in growth signals; the Met gene can be over expressed; or the Met receptor can become overactive by mutations [79], [78]. The HGFA-HGF-Met pathway has been implicated in many forms of cancer and it is one of the most frequently dysregulated pathways in human cancer (Table 1.4) [80]. The HGFA-HGF-Met pathway is an appealing drug development target as it is involved in so many cancers and regulates such a wide span of functions.

It has been shown that Met mediated cancer cell transformation is dependent upon HGF activation of Met, making the HGF-Met interaction an attractive point for drug development [81], [82]. Many different ways to target the HGFA-HGF-Met signalling cascade are being explored: small molecule tyrosine kinase inhibitors [83], [84]; gene therapy [85]; monoclonal antibodies targeting both HGF and Met [80], [81], [86]; injecting HGF-NK1 [87]; and injecting Met [79]. Small molecule tyrosine kinase inhibitors and antibodies towards Met seem to be the most common approaches, often used in combination with other therapies [80]. The activation of HGF by HGFA and matriptase is also being targeted for drug development by HGF activation inhibitors and monoclonal antibodies [88].

Table 1.4 The involvement of HGF and Met in different types of cancer and supporting data. Y: Yes, N: No/Not yet. Taken from [89].

Category	Cancer Type	HGF/SF expression	Met expression	Poor Prognosis	Mutation of Met	In vitro studies	Animal model	Therapeutic Development
Carcinomas								Y
	Bladder	Y	Y	Y	N	Y	Y	Y
	Breast	Y	Y	Y	Y	Y	Y	Y
	Cervical	Y	Y	Y	N	Y	N	Y
	Cholangiocarcinoma	Y	Y	N	N	Y	Y	Y
	Colorectal	Y	Y	Y	N	Y	Y	Y
	Endometrial	Y	Y	N	N	Y	N	Y
	Esophogeal	Y	Y	Y	N	Y	N	Y
	Gastric	Y	Y	Y	Y	Y	Y	Y
	Head and Neck	Y	Y	Y	Y	Y	Y	Y
	Kidney	Y	Y	Y	Y	Y	Y	Y
	Liver	Y	Y	Y	Y	Y	Y	Y
	Lung	Y	Y	Y	Y	Y	Y	Y
	Nasopharyngeal	Y	Y	Y	N	Y	N	N
	Ovarian	Y	Y	Y	Y	Y	Y	N
	Pancreas/Gall Bladder	Y	Y	N	N	Y	Y	Y
	Prostate	Y	Y	Y	N	Y	Y	Y
	Thyroid	Y	Y	Y	Y	Y	N	Y
Musculoskeletal sarcomas		Y	Y			Y		
	Osteosarcoma	Y	Y	N	Y	Y	Y	Y
	Rhabdomyosarcoma	Y	Y	Y	N	Y	Y	Y
	Synovial Sarcoma	Y	Y	Y	N	Y	N	N
Soft tissue sarcomas					Y		Y	
	Kaposi's Sarcoma	Y	Y	N	N	Y	Y	N
	Leiomyosarcoma	Y	Y	Y	N	Y	Y	Y
	MFH/Fibrosarcoma	Y	Y	N	N	Y	Y	Y
Hematopoietic Malignancies								
	Acute Myelogenous Leukemia	Y	N	N	N	Y	N	N
	Adult T Cell Leukemia	Y	Y	Y	N	Y	N	Y
	Chronic Myeloid Leukemia	Y	N	N	N	N	N	N
	Lymphomas	Y	Y	Y	Y	Y	Y	Y
	Multiple Myeloma	Y	Y	Y	N	Y	Y	Y
Other Neoplasms		Y						
	Glioblastomas/Astrocytomas	Y	Y	Y	Y	Y	Y	Y
	Melanoma	Y	Y	Y	N	Y	Y	Y
	Mesothelioma	Y	Y	Y	N	Y	N	Y
	Wilms' Tumour	Y	Y	N	N	Y	Y	Y

HGF-NK1 has been described both as a Met antagonist [90] and a partial agonist [91]. For therapeutic applications, HGF-NK1 and HGF-NK4 are being developed as antagonists [87]. Interestingly, it has been shown that HGF-NK4 is able to inhibit angiogenesis independently of its HGF/Met antagonism. It has been suggested that HGF-NK4 achieves this by binding perlecan, competing with fibronectin and other perlecan binding growth factors necessary for angiogenesis [92]. It is likely that HGF-NK1 exhibits similar properties.

A relatively novel class of drugs targets the interaction between GAGs and proteins involved in cancer, such as GAG binding growth factors and heparanase [93]. The wide importance of protein-GAG interactions in cancer makes them appealing targets for drug development [94]. What is also appealing is that one GAG mimetic could work on multiple targets, for example HGF, FGF, VEGF, PDGF and heparanase, potentially making it a very effective class of drugs that can be used in combination with other more targeted treatments. These GAG mimetic drugs structurally resemble GAGs and are often highly negatively charged. Importantly, they have no or very low anti-coagulant activity [95].

Unfortunately, the use of a promiscuous GAG mimetic has the potential to cause severe side effects as GAGs are involved in many important physiological processes. Another problem might be the bioavailability of *per os* tablets, as many GAG mimetics composed of sugars will be degraded in the gastrointestinal tract and during first-pass metabolism. A GAG mimetic might therefore have to be administered in a manner similar to heparin anticoagulant therapy: intravenously or subcutaneously.

The GAG mimetics developed to date range from sulfated peptides [96], [97], chemically oversulfated oligosaccharides [38], [98-100], sulfated carrageenan sugars from algae [101-105], modified GAGs [4], [106], [107], specifically synthesised heparin like oligosaccharides [30] to small synthetic molecules [108]. Many of the GAG mimetics show promising *in vivo* and *in vitro* results, but relatively few are in

clinical trials. Heparin anticoagulant therapy has been shown to decrease cancer mortality in several clinical trials [109], [110], strongly supporting the idea to use GAG mimetics as anti cancer agents [95], [111].

The development of a HGF specific GAG sequence similar to the FGF specific GAG sequence [30] could potentially be aided by the NK1-heparin complex crystal structure [14]. However, it has to be kept in mind that GAG oligosaccharides longer than DP4 can work as co-activators for HGF [112], something that has to be avoided in drug design. However, it is questionable if the HGF-GAG interaction is specific enough and it is unlikely to be as specific as the ATIII binding pentasaccharide, which seems to be a particularly specific protein-GAG interaction.

An example of a successful GAG mimetic in development is Muparfostat, originally called PI-88 or phosphomannopentaose, developed by Progen Pharma as a general GAG mimetic (Figure 1.10). The drug has heparanase inhibitory properties, binds FGF and VEGF with high affinities and its main mode of action is that it is anti-angiogenic [5-8]. Muparfostat showed promising results in phase II clinical trials and is now in phase III for advanced melanoma and post resection liver cancer. New chemical entity analogues of Muparfostat called the PG500 series have been synthesised to improve pharmacokinetics and decrease anti-coagulant activity by attaching different lipophilic groups to the reducing end [113]. Muparfostat shows that GAG mimetic drugs are feasible.

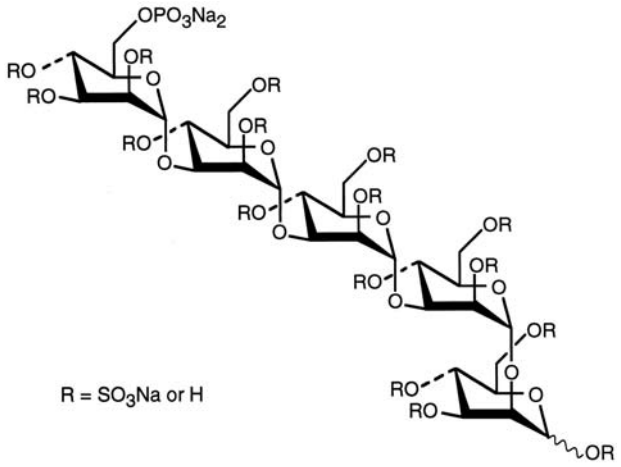


Figure 1.10 The anti-cancer drug Muparfostat is composed of five sulfated mannose units. Taken from [7].

As explained in section 1.3.6, GAG-protein interactions are also important players in regulating inflammation, which is tightly linked to cancer progression and there are also considerable possibilities for targeting protein-GAG interactions for anti-inflammatory drug development.

1.2.7 HGF Research

As HGF was first purified, cloned and expressed in the early 1980s [73], [74], HGF research was initially relatively slow. When Met was discovered in 1984 [76] and shown to be the receptor for HGF [75] in 1991, research into HGF/Met and its physiological roles increased quickly. However, similar to GAG research, the pace of HGF research has dropped somewhat in the new millennium compared to PubMed indexed research.

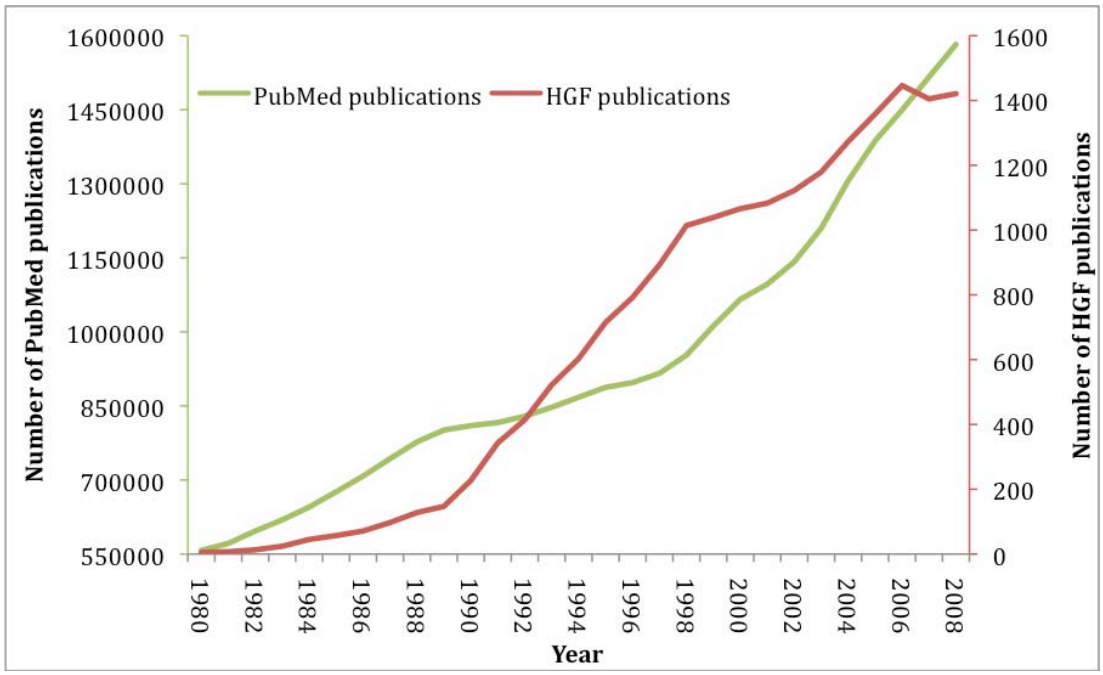


Figure 1.11 The graph shows number of total PubMed publications on the left axis in green and number of HGF publications in red on the right axis sorted per year. A PubMed search was performed for each year with the search term "(((((HGF) OR HGF/SF) OR "Hepatocyte growth factor") OR "Scatter factor") OR Hepatopoeitin-A)".

1.3 Glycosaminoglycans

The extracellular matrix consists largely of complex protein linked carbohydrates called glycosaminoglycans (GAGs) that bind water and work as a scaffold [3]. Although GAGs might not make up a high percentage of the extracellular matrix weightvise, they form a hydrated gel that occupies a very large volume relative to their mass [18]. These highly elongated and negatively charged GAGs interact with a number of proteins such as growth factors, enzymes, complement proteins and cytokines. In recent years it has become clear that GAGs are not just hydrating/scaffolding molecules, but also play a central role in homeostasis [109], host-pathogen interactions [114], development [115] and cell signalling [116].

1.3.1 Proteoglycans

GAGs often occur linked to proteins, forming proteoglycans. There are many different types of proteoglycans [117] and they can be classified according to the kind of GAG they carry, for example HSPG being a heparan sulfate (HS) proteoglycan. Proteoglycans can also be classified depending on what protein class they belong to or their physiological location. The cell-surface classes of proteoglycans that are linked to the cell-membrane are the glypicans [118], [119] and syndecans [120]. These two classes of proteoglycans carry HS chains (although hybrid forms containing both HS and CS (chondroitin sulfate) have been found in some tissues) and are likely to be the most relevant for the HGF-GAG interaction [119]. Glypicans are linked to the cell membrane by a GPI anchor whereas syndecans are transmembrane proteins. Syndecans typically carry 3-5 GAG chains and glypicans typically carry 2-3 GAG chains. Interestingly, the glypican GAG chains are attached close to the transmembrane region of the protein sequence, placing the GAG chains closer to the cell membrane than the main extracellular parts of the protein. This probably allows for GAG-binding signalling molecules to bind their cell surface target receptor easier. Syndecans are also found free in the extracellular matrix as a protease site in the ectodomain allows for shedding of the whole proteoglycan [121]. Shedding of the ectodomain is likely important in signalling regulation [122]. Other classes of proteoglycans such as the perlecans that present in the extracellular matrix in basal membranes have also been shown to interact with HGF [123].

1.3.2 Primary Structure of Glycosaminoglycans

GAGs are divided into different groups depending on their repeating disaccharide structure and acetylation/sulfation pattern [124] (Figure 1.12 and Figure 1.13). Most GAGs are linked via a glucuronic acid monosaccharide, two galactose monosaccharides and one xylose monosaccharide to a serine or threonine amino acid residue on proteins to form proteoglycans.

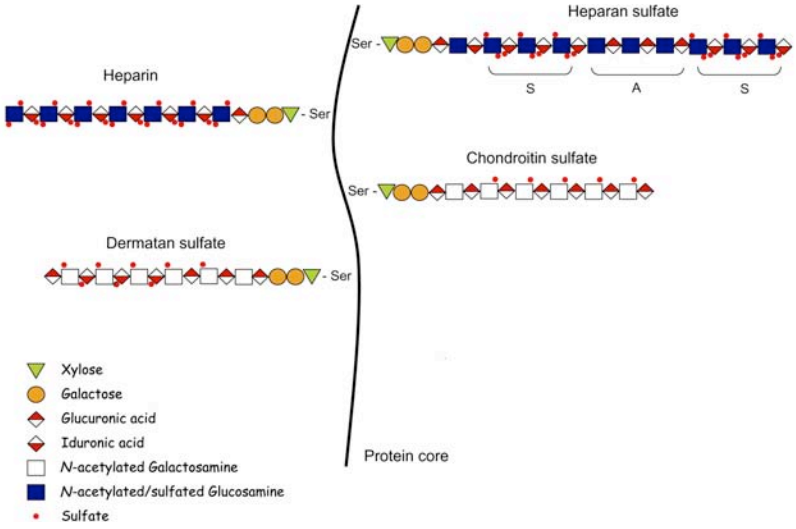


Figure 1.12 Overview of the different GAG structures, showing their sugar constituents and sulfation patterns. The GAG chains are shown as being attached to a protein molecule, forming proteoglycans. The S and A regions on HS indicated N-sulfated (NS) and N-acetylated (NA) regions. Taken from [124].

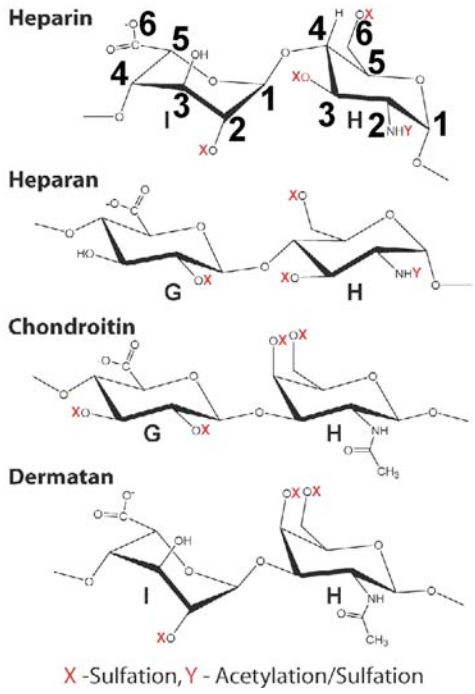


Figure 1.13 Main disaccharide units in the different types of GAGs, H: Hexoseamine, G: Glucuronic acid, I: Iduronic acid. Carbon backbone numbering is shown for heparin. Note that the disaccharides are somewhat interchangeable, for example heparin-like disaccharides can be present in HS. Modified from [125].

The different classes of GAGs are heparin, heparan sulfate, chondroitin sulfate, dermatan sulfate (DS), hyaluronic acid and keratan sulfate (the two latter have been left out in Figure 1.12 and Figure 1.13 as they are not relevant for this project). There are also GAG subclasses such as chondroitin A, C and E [116] and reports about CS/DS hybrid GAGs [126], so the classification is clearly flexible and not always straightforward.

Heparan sulfate is mainly built up of glucuronic acid (GlcA) that can be 2-O-sulfated and is 1-4 linked to a glucosamine (GlcN) that can be N-acetylated (GlcNAc) or N-sulfated. If the GlcA is N-sulfated, it can undergo epimerisation to iduronic acid, followed by O-sulfation at C-2 and C-6 of the GlcNS, producing a heparin like disaccharide. The C-3 position of HS disaccharides can be O-sulfated, usually in conjunction with N, C-2, and C-6 –sulfations, but only 0.15% of the disaccharides in HS are tetrasulfated [127]. Heparin consists mainly of iduronic acid (IdoA) that is 2-O-sulfated and 1-4 linked to a glucosamine that is usually N-sulfated (GlcNS) on the C-2 position and O-sulfated on the C6 position. The rare 3-O-sulfate that exists in HS also exists in heparin. The difference between heparin and HS is that the latter is more heterogeneous, contains extensive monosulfated domains and that HS also contains unsulfated disaccharides.

Chondroitin sulfate (CS) consists of glucuronic acid 1-3 linked to an N-acetylated galactosamine (GalNAc). If the GalNAc is sulfated on the four position the species is classified as CSA, if the GalNAc is 6-sulfated the species is classified as CSC and if the species is both 4 and 6 –sulfated it is classified as CSE. Dermatan sulfate (DS) consists mainly of iduronic acid 1-3 linked to an N-acetylated galactosamine, which can be sulfated on the four or six positions of the GalNAc.

Heparin is probably one of the most well known GAGs as it is a common anticoagulant drug [109]. Chondroitin is also relatively well known for working as a lubricant in joints. This project focuses mainly on heparin, CS and DS.

1.3.3 Biosynthesis of Glycosaminoglycans

The heterogeneity of HS and other GAGs arises during the biosynthesis in the Golgi apparatus and the action of post-synthesis GAG modifying enzymes such as sulfatases that can modify the GAG motifs in the Golgi apparatus and on the cell surface [128]. GAG chain biosynthesis initiates after a xylose has been transferred to a translated protein chain that is destined to become a proteoglycan (Figure 1.14). The xylose is attached to a Ser-Gly dipeptide with one or several neighbouring acidic amino acids. After the xylose has been conjugated, two galactose sugar units are attached to it by galactosyl transferase I and II. Next, a glucuronic acid is appended by glucuronosyltransferase I. The GlcA-Gal-Gal-Xyl-[Ser] tetrasaccharide is the common core protein linkage sequence for heparin, HS, CS and DS. The linkage galactoses can be both sulfated (CS) and phosphorylated (CS, HS, heparin). The enzymes involved in the biosynthesis of this linker are vital for GAG chain initiation and without them there would be no GAG chains at all.

Interestingly, the GlcNAc/GalNAc transferases not only recognize the tetrasaccharide, but also the amino acid motifs to which the tetrasaccharide is attached and both the tetrasaccharide sequence and the amino acid motif determine whether the synthesised GAG type will be of CS/DS or heparin/HS type.

After the linkage tetrasaccharide has been attached and an α-linked GlcNAc has earmarked it for HS synthesis, the chain is elongated by alternating additions of GlcA and GlcNAc units by enzymes from the EXT gene family working synergistically with each other [129], [130]. The sugar units are transferred from UDP-GlcA or UDP-GlcNAc nucleotides.

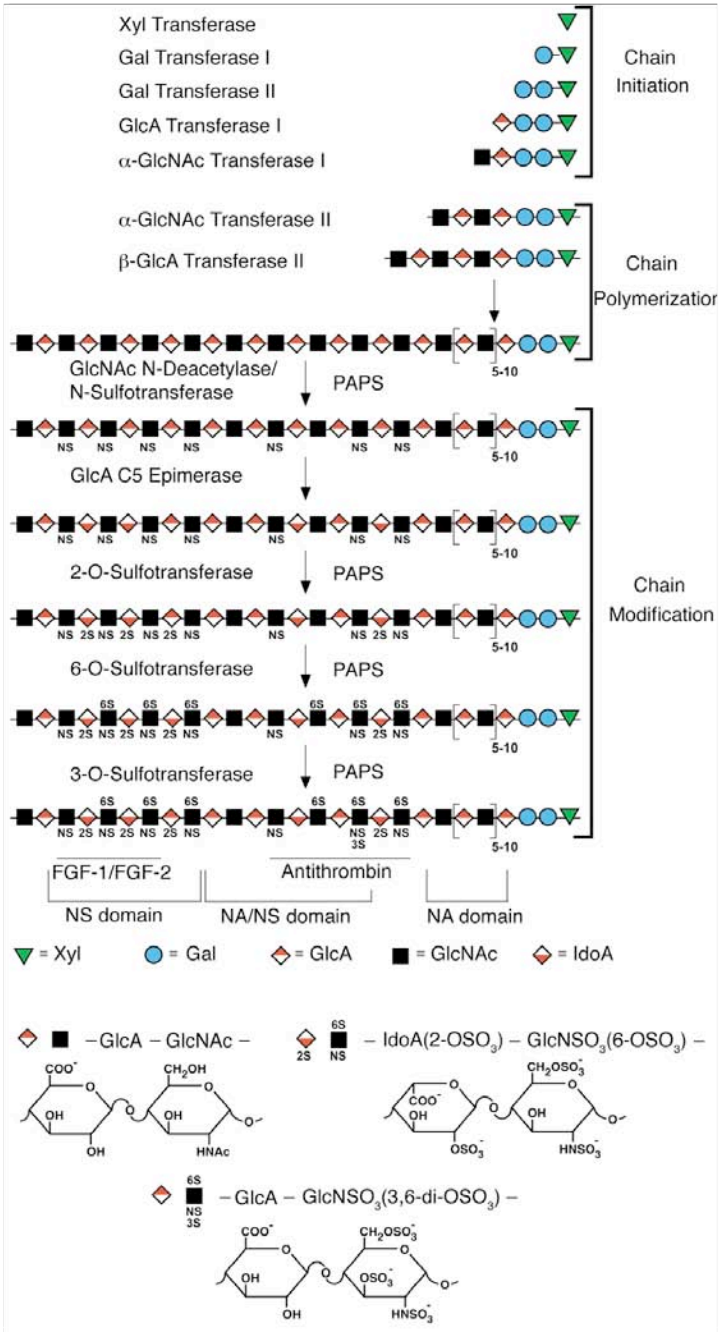


Figure 1.14 The sequential enzymatic steps in heparin/HS biosynthesis. NA: N-Acetylated, NS: N-Sulfated, 2S: 2-Sulfated. Taken from [121].

After the linkage tetrasaccharide has been assembled, the chain is committed to HS elongation by the addition of an α-linked GlcNAc residue by αGlcNAc transferase, whereas the addition of a β-linked GalNAc commits the chain to CS/DS elongation.

As the chain is being elongated and finished, sulfotransferases and epimerases start to modify the HS chain. Some of the modifications take place in sequential order and others can take place unsystematically (Figure 1.14):

a) GlcNAc N-deacetylase/N-sulfotransferase (NDST) removes the N-acetyl group of GlcNAc units and replaces it with a sulfate group. The NDST enzymes catalyse both reactions and the sulfate group is supplied by a PAPS cofactor. PAPS is an universal sulfate donor in biochemical reactions, it is somewhat similar in structure to ATP in that it contains an adenosine, but it donates a sulfate instead of a phosphate. The different NDST enzymes are expressed at different levels in different tissues and have varied deacetylation and sulfation efficiencies, for example NDST3 is an efficient deacetylater whereas NDST4 exhibits good sulfotransferase activity. The enzymes are likely involved in the regulation of HS sequences, for example NDST1 and NDST2 are widely expressed, whereas NDST3 and NDST4 have only been identified in adult brain and fetal tissues. The varied efficiencies in deacetylations/sulfations could explain why 0.7–4 % of deacetylated glucosamines in different HS and heparin preparations lacks an N-sulfate [131]. The de-N-acetylated disaccharides are often concentrated in the protein linkage region, but in aortic HS they also occur more peripherally, in general the free amines are located within the NA domains, or in transition NA/NS domains. The de-N-acetylated disaccharides contain free amines have been shown to be important in some protein-GAG interactions [132], [133]. As further explained in section 1.3.4, the NS regions are interspaced between NA regions in HS. How the overall structure of these NA/NS domains is determined is unclear, but it is thought that the NDST enzymes play a crucial role in the regulation [134]. The replacement of the N-acetyl group by a sulfate is the first modification in the GAG biosynthesis and is required for several subsequent modifications. The NDST enzymes are inhibited by O-sulfates on the disaccharide unit, which is an important regulatory function it helps the biosynthesis take place in a specific order.

b) After the GlcNAc residues have been N-sulfated, the reducing end neighbouring GlcA can be epimerised to IdoA by C-5 epimerase. To date, only one C-5 epimerase has been identified. The same enzyme epimerises heparin and HS chains [135] and interestingly, the epimerisation reaction is irreversible *in vivo*, but not *in vitro* [136]. The enzyme does not react with any GlcA residues that are O-sulfated or that are next to an O-sulfated glucosamine residue. This suggests that epimerisation happens after the first N-deacetylation/N-sulfation but before any O-sulfation reactions. Glucuronic acid epimerisation is regulated by: 1) the extent of N-sulfation, 2) the level of C-5 epimerase expression, 3) the extent of O-sulfation. IdoA is a very common residue in protein binding GAG motifs because it can adopt more conformations than GlcA as further discussed below.

c) 2-O-sulfotransferase is thought to act at the same time as the NDST and C-5 epimerase enzymes. Using PAPS as a cofactor, 2-O-sulfotransferase transfers a sulfate to the 2-O position of both IdoA and GlcA, but acts more efficiently on IdoA and thus more IdoA units are sulfated compared to GlcA. As expected, 2-O sulfation is more common in NS domains, but it is also present in NA domains, for example human cerebral cortex has a relatively high 2-O-sulfated GlcA content [137]. As 2-O-sulfotransferase performs its actions at the same time as NDST and C-5 epimerase and the resulting 2-O-sulfate group inhibits the two latter enzymes, 2-O-sulfation is a regulatory step. To date, only one 2-O-sulfotransferase has been found.

d) The three members in the 6-O-sulfotransferase class of enzymes (6-OST-1, 6-OST-2 and 6-OST-3) exhibit different substrate specificities, but they all catalyse the same reaction: addition of a 6-O-sulfate to glucosamine residues [138]. 6-O-sulfates seem to be more common in NS domains. 6-OST-2 exists in two different splice forms with different tissue expression patterns [139]. 6-OST-1 is expressed in liver, 6-OST-2 is expressed mainly in brain and spleen and 6-OST-3 is expressed ubiquitously. It is thought that 6-O-sulfation typically happens in the later stages of heparin/HS synthesis, although some flexibility probably exists.

e) Both 3-O-sulfation and 6-O-sulfation can be the final step in heparin/HS biosynthesis [140], although 3-O-sulfation is often thought to be the last step that only modifies a few residues per chain [141]. It is the most uncommon sulfation in heparin and heparan sulfate and is important in several protein binding motifs, for example for herpes simplex virus glycoprotein D, FGF-7 and antithrombin III. Six different 3-O-sulfotransferases with different substrate specificities and tissue expression patterns have been identified to date. As can be seen in Figure 1.15 and Table 1.5, the several isoforms of 3-O-sulfotransferases allows for a tight regulation of specific GAG sequence motifs.

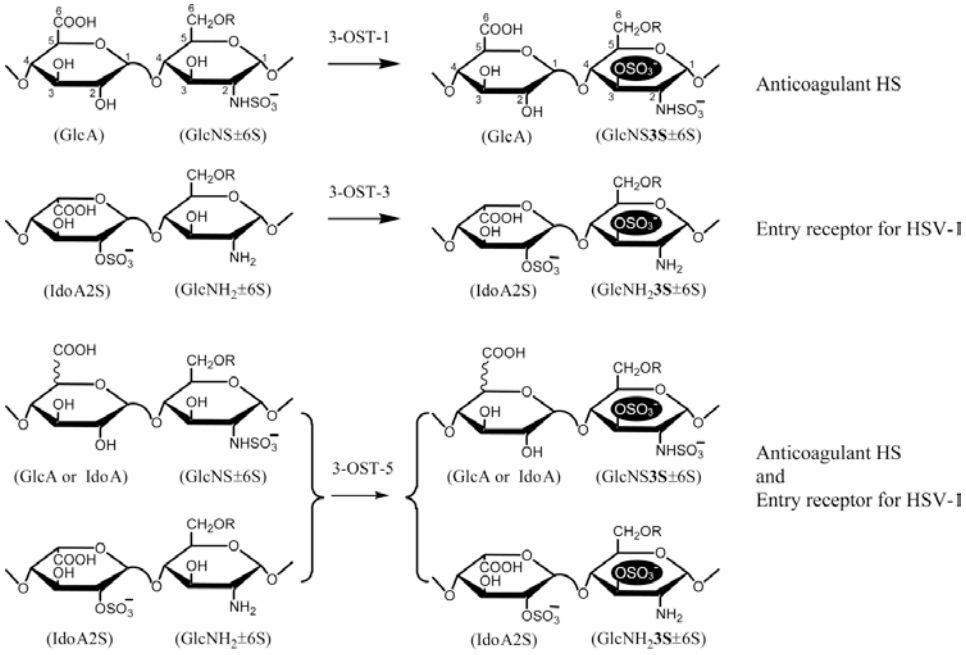


Figure 1.15 HS/heparin disaccharide specificity for the different 3-O-sulfotransferases, the products and examples of product importance. Taken from [142].

Table 1.5 The table summarises the different 3-O-sulfotransferase disaccharide specificities. Summarised from [141-147].

Enzyme	**Target**
3OST-1	GlcA-GlcNS±6S
3OST-2	GlcA2S/IdoA2S-GlcNS
3OST-3A and -3B	IdoA2S-GlcNH2±6S
3OST-4	?
3OST-5	HexA±2S-GlcNH2±S6S
3OST-6	IdoA2S-GlcNH2±6S (same as 3OST)

After the GAG chains have been synthesised and the proteoglycans presented on the cell surface, post-synthesis modification can take place by the removal of sulfates by sulfatases [148]. QSulf1 and QSulf2 are sulfatases that remove 6-O sulfates from HS both in the Golgi apparatus as the GAGs are being synthesised and post-synthesis, extracellularly [149]. The enzymes are important in cell signalling, for example it has been shown that they promote Wnt signalling, but inhibit FGF signalling [150]. These sulfatases are special in the way that by removing sulfates they translate the GAG sequence information into a biological event, setting them apart from other sulfatases that are often involved in metabolism [151]. Both QSulf1 and QSulf2 have the same substrate disaccharide specificity for heparan sulfate NS domain disaccharides, namely UA±2S-GlcNS6S (a disaccharide consisting of an uronic acid that is optionally 2-O-sulfated, followed by a galactosamine that is N-sulfated and 6-sulfated) [152]. The two enzymes are not covalently attached to the cell surface as they can be washed off with high salt buffer.

As follows from the above discussion, there is a great deal of flexibility in GAG biosynthesis machinery. Although some reactions take place sequentially, many reactions take place in a random order and some enzymes work at relatively random

sites, but through the careful orchestration of isoenzyme expression in different tissues, specific motifs can be synthesised as needed. Most of the GAG biosynthesis takes place in the Golgi apparatus, where it has been shown that many of the involved enzymes assemble in a big complex, called the GAGosome [141].

Much of what is known about the physiological roles of GAGs, particularly in development, has been learned from knocking out particular enzymes in model organisms such as mice [153]. Although some enzymes are vital, there is often a surprising flexibility and redundancy in the GAG synthesis machinery and similar GAG motifs can be synthesised even when isoenzymes have been knocked out. This is likely because other isoenzymes can step in for the knocked out enzyme. In some enzyme knockout models, modified GAGs are produced that still can fulfil their physiological roles. This compensatory effect of one GAG motif for another emphasises that although GAG binding proteins usually prefer a specific motif, the interactions are promiscuous and the proteins can bind alternative GAG motifs with a lower affinity and still function.

Some of the GAG biosynthesis enzymes are involved in diseases, for example EXT1 is a tumour suppressor that has been implicated in cancers and hereditary multiple exostoses, a bone disease [154].

1.3.4 The Heterogeneity of Glycosaminoglycans

Referring back to Figure 1.12, it can be seen that heparan sulfate contains interspaced highly and lowly sulfated regions [121]. The lowly sulfated regions are called NA domains as they are N-acetylated and the highly sulfated regions are called NS regions as they are N-sulfated. To further add to the complexity, HS also contains NA/NS domains, which is a mix of NA and NS disaccharides. The length of a NS-domain is roughly 2-9 disaccharides, being interspaced by 16-18 disaccharides of NA and NA/NS domains [155], consequently the NA domains take up the largest part of the HS chains with the shorter NA/NS domains bordering to the NS domains

(Figure 1.16). However, the degree of HS sulfation and how tightly spaced the NS domains are seems to vary between and within different organs [156], [155]. For example, rat liver HS is much richer in NS domains and its chains are shorter than rat skin fibroblast HS.

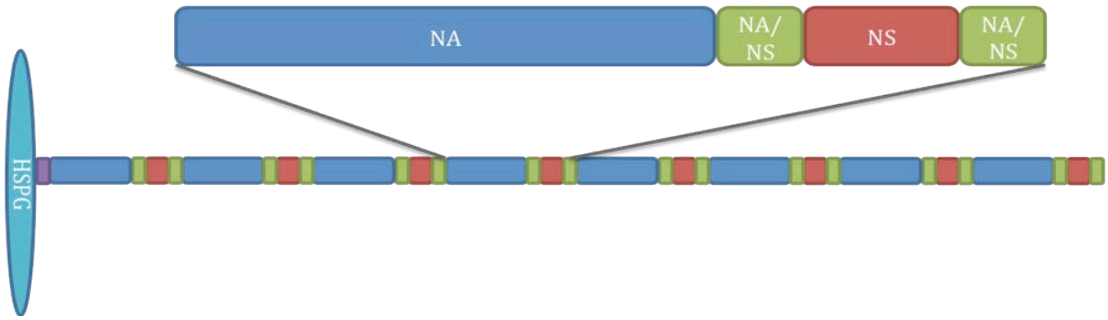


Figure 1.16 Domain structure of a heparan sulfate. The purple unit is the tetrasaccharide protein-GAG linkage.

It is not only the overall NA/NS domain structure that varies between tissues, but sugar units and sulfation patterns also differ between tissues on a very specific level [157]. The HS disaccharide fingerprint has also been found to change with age [158] and different diseases [159]. All these differences in HS sequences will be recognised differently by signalling and regulatory proteins and are ways of regulating signalling pathways in physiological processes, disease and during embryogenesis [115]. The disaccharide composition, expression levels of GAG modifying enzymes and sulfation patterns of HSPGs in different cancers have been identified as possible bio-markers for diseases. For example an up-regulation of the 3-OST-1 sulfotransferase and an increase in 3-O-sulfation in hepatocellular cancer was identified by RT-PCR and monoclonal antibodies recognising the 3-O-sulfate epitope [160].

It is clear that the heterogeneity of GAGs carries information in a manner similar to DNA (although GAGs are not replicated or translated) and plays a central role in homeostasis and regulation of signalling pathways. It is well known that CS and DS are less heterogeneous than heparin and HS and physiological regulation has often

been pinpointed to the latter two GAGs [124], [125], [161], [162], [66], [163]. Although variance exists in DS and CS GAGs too and this heterogeneity is also likely to be important for regulating biological activity, for example signalling pathways [126], [164], [40], [165].

The highly sulfated NS regions of HS very much resemble the heparin structure and it is thought that proteins preferably interact with these highly sulfated IdoA containing regions in HS. Heparan sulfate is found on the cell surface and on matrix proteoglycans and is thought to be the natural *in vivo* ligand for HGF and most other heparin binding proteins such as the GAG binding growth factors FGF (Fibroblast Growth Factor, a family of heparin binding growth factors involved in angiogenesis, wound healing, and embryonic development) and VEGF (Vascular Endothelial Growth Factor, a family of growth factors produced by cells that are exposed to low oxygen levels that stimulates angiogenesis). However, heparin is often used in place of heparan sulfate in experiments as heparin is readily available, has a more homogeneous sulfation pattern and has a more or less identical structure to the HS NS domains [166]. Heparin mimics the highly sulfated NS regions of HS, which are usually the protein binding regions that are of most interest. This is also the argument of using heparin instead of heparan sulfate in this thesis.

The golden grail of protein-GAG interaction research is to understand how proteins recognise a specific GAG motif. Although there has been some achievement in the field, such as the identification of the ATIII binding pentasaccharide, the hunt for the exact GAG epitope of a protein-GAG interaction is not trivial and complexity quickly increases with increasing GAG length [167]. To date 23 different disaccharide structures have been identified in heparin and HS and considering that a typical binding motif might be a hexasaccharide to a decasaccharide there are 23^3 to 23^5 possibilities to combine the disaccharides (Figure 1.17).

There are several tools available to identify the binding motifs, for example affinity chromatography, microarrays, SPR and ITC. There are also several tools available to identify the disaccharide composition of GAGs, for example disaccharide analysis, mass spectrometry and NMR. Disaccharide analysis is a powerful technique that can elucidate both disaccharide contents and sequence [168], [169]. NMR also has the power to elucidate the complete sequence, but demands relatively large quantities of material that sometimes can be hard to achieve. One might think that with these powerful tools, it should be straightforward to identify binding motifs, but the problem is often complicated by promiscuous binding and by the fact that some proteins have similar affinities for several motifs. Furthermore, like in the ATIII case, the binding motif can contain an uncommon 3-O-sulfate that is only present in small quantities.

Heparin and other GAGs obtained from different sources and batches differ in sugar units, sulfation levels, patterns, sequences and chain lengths [170], [171], [156]. If the GAGs are not further processed and purified or thoroughly characterised, experimental results may well lead to different results in a series of analogous experiments.

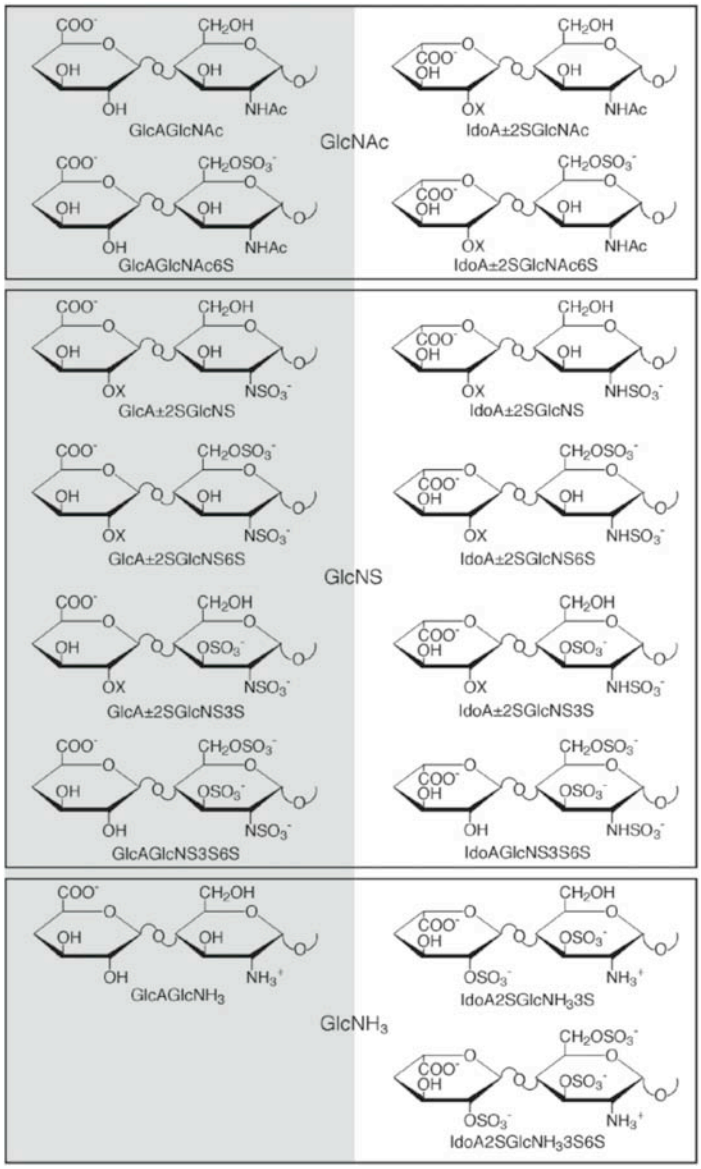


Figure 1.17 Twenty three identified disaccharide structures in heparin and HS. GlcA-containing disaccharides are shown on the left and IdoA-containing disaccharides are shown on the right. X= H or SO_3^-. Taken from [128].

1.3.5 Secondary structure of Glycosaminoglycans

Figure 1.18 shows the NMR structure of heparin and as can be seen, heparin adopts a straight helical DNA-like structure in solution without any tertiary structure [172].

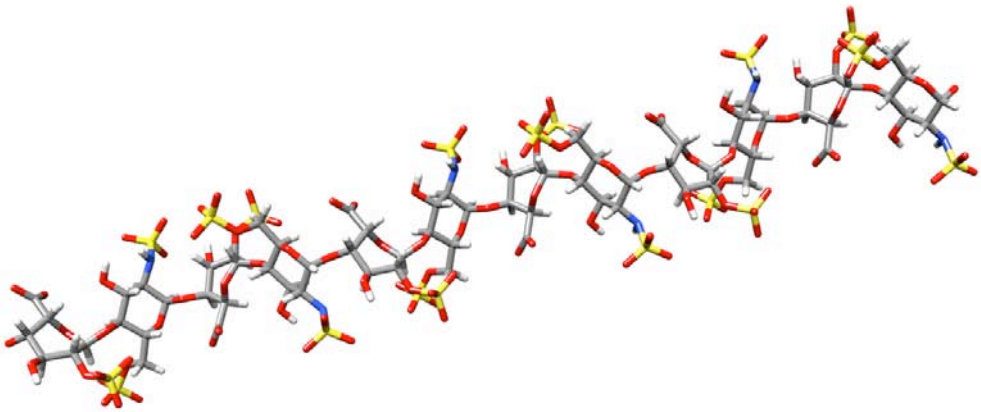


Figure 1.18 NMR structure of heparin [173]. The 1hpn PDB file with the IdoA residues in a $^{1}C_{4}$ conformation was rendered in UCSF Chimera 1.4.

Different GAGs can adopt different helical shapes [167], where the most important parameters that determine the helical structure are the glycosidic linkage torsion angles and the ring conformation of the sugar units. The glucosamine and galactosamine residues adopt a $^{4}C_{1}$ conformation, which is also the preferred conformation of glucuronic acid (Figure 1.19). Iduronic acid is the most flexible GAG component; as the energy difference between the $^{1}C_{4}$ and $^{2}S_{0}$ conformations is very low, it can adopt both conformations. The energy difference between the conformations is too high for the other sugar residues. As iduronic acid is more flexible than glucuronic acid, this also makes the NS domains in HS relatively flexible and the NA domains relatively rigid. The preferred conformation of IdoA depends on its 2-O substitution and on the substitution pattern of the neighbouring GlcN.

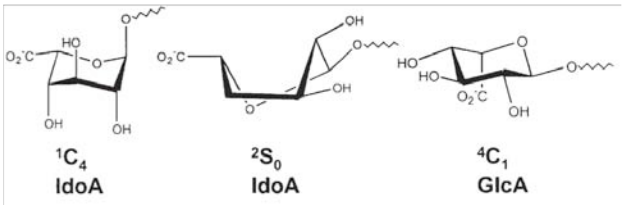


Figure 1.19 The most common conformations of the GlcA and IdoA residues. Taken from [174].

1.3.6 Physiological Roles of Glycosaminoglycans

GAGs have many physiological roles, one of the classic examples being regulation of blood clotting. In this case, antithrombin-III (AT-III), a serine protease inhibitor, specifically recognizes a HS pentasaccharide fragment containing a 3-O sulfate. When AT-III has recognized and bound the pentasaccharide, its serine protease inhibiting activity drastically increases and by inhibiting proteases in the blood-clotting cascade, homeostasis is maintained [174]. This interaction is one of the most common examples of a highly specific protein-GAG interaction.

Interestingly, patients on LMWH (Low Molecular Weight Heparin) anticoagulant therapy have reduced incidence of cancer [110]. This is likely due to inhibition of heparin binding growth factors such as HGF, FGF and VEGF, but could also be due to heparin inhibiting heparanase and selectins that are vital for cancer cell metastasis [95], [175]. Inhibition of heparanase and selectin mediated extravasation has also been suggested to be the physiological role of heparin as an anti-inflammatory agent released from mast cells [176]. As can be seen in Figure 1.20, selectins on immune cells interact with proteoglycans on the endothelium and allow immune and cancer cells in blood vessels to attach to the walls and extravasate to the inflammation site, or in the case of cancer cells, metastase. This points to the importance of protein-GAG interactions not only in cancer, but also in inflammation, two important pathological processes that are being targeted for drug development by GAG-mimetics [4].

Figure 1.20 points out many of the important protein-GAG interactions that are targetable for cancer and inflammatory drug development. The start of an inflammatory response can for example be that tissue damage activates macrophages to release chemokines and other cytokines. Essentially all chemokines are able to interact with GAGs as they are diffusing through the extracellular matrix [177]. The cytokines and chemokines carry signals to the endothelial cells to display proteoglycans such as syndecans and selectins, enabling circulating cells to bind to the endothelium. Once the immune cells have bound to the endothelial cells and

started to extravasate, heparanase and proteases are released by the immune cells to degrade the basal membrane and extracellular matrix so that the cells can migrate [178]. As the extracellular matrix is degraded, immobilised growth factors that have been bound to the previously intact proteoglycans are freed. These released growth factors (VEGF, FGF and HGF for example) can increase angiogenesis, supporting inflammation or cancer growth. For example, it has been shown that FGF2 is stored in a dormant form bound to syndecans that is cleaved by heparanase as a mitogenic complex [4]. The release of heparanase itself is also important for angiogenesis, as the basal membrane extracellular matrix needs to be degraded and remodelled for angiogenesis [179].

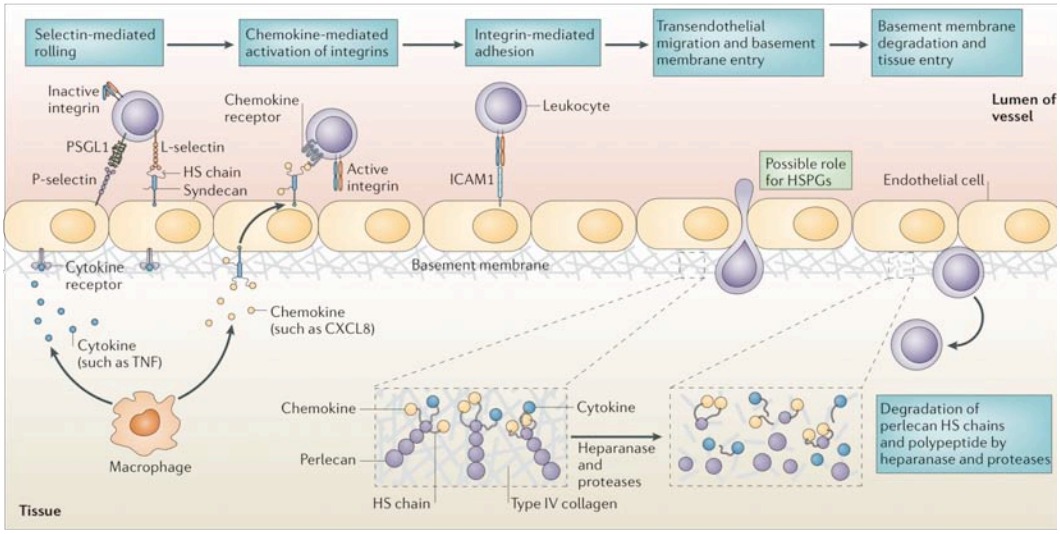


Figure 1.20 The importance of protein-GAG interactions in the migration of immune cells from blood vessels. Taken from [178].

Heparanase has been shown to be overexpressed in many types of cancers and in tumours it confers an accelerated growth and invasive phenotype [180]. One of the main roles of heparanase in cancer is likely to free immobilised HS-binding growth factors through its enzymatic activity [4], [181]. The mobilised growth factors can promote angiogenesis, cell proliferation and survival. The oligosaccharides that are produced when heparanase digests HS can work as co-receptors for the released growth factors. Through degrading and remodelling the extracellular matrix, heparanase contributes to tissue remodelling and cancer cell invasion. Independently

of its enzymatic activity, it has been shown that heparanase stimulates Akt-dependent cell invasion and migration [182]. Interestingly, overexpression of heparanase has also been shown to increase tissue factor activity, which in a second step can increase the levels of HGFA (Hepatocyte Growth Factor Activator) and in a third step increase the levels of active HGF, contributing to further cancer progression (Figure 1.8) [181]. It is possible that LMWH therapy inhibits heparanase or competes with its natural target heparan sulfate. This would decrease the effects that heparanase has on cancer development and progression [179].

Heparin is only found in mast cell rich tissues such as the trachea and intestinal mucosa. It is synthesised as a proteoglycan called serglycin, but is cleaved from its core protein before it is stored in granules of immune cells such as mast cells and basophiles. The heparin is secreted as a pure glycosaminoglycan chain after IgE stimuli [183]. Serglycin typically has 10-15 GAG chains attached, which can also be HS, CS and DS, depending on tissue and cell type [184]. The role of heparin *in vivo* is still somewhat unclear, but it has been shown that it is vital for the storage and release of proteases that regulate inflammatory responses from mast cell granula. It has therefore, in conjunction with other evidence been suggested that heparin has an anti-inflammatory effect as discussed above [185]. It has been shown that inhaled heparin prevents exercise-induced asthma [186]. The action of heparin in this case is likely a secondary effect related to a decrease of smooth muscle constriction mediators such as histamine from mast cells. GAG-protein interactions are important in many steps of the inflammatory process and it is not surprising that exogenous heparin can modify the process. On a molecular level, heparin is likely to bind to selectins and inhibit the interaction with their natural sialyl Lewis^x ligand [176], [187]. Selectins are displayed on the blood vessel endothelium at the location of inflammation. If heparin binds to the selectins, immune cells will not be able to extravasate to the inflammation site and the inflammatory response will be suppressed (Figure 1.20). A big effort is currently going into identifying GAG mimetics without anticoagulant properties but with anti-inflammatory and anticancer properties.

1.3.7 Protein-Glycosaminoglycan Interactions

The ionic interactions of sulfates and carboxyls with positively charged lysines and arginines are usually the most important protein-GAG interaction points and thus the sulfation positions/pattern of the GAG is the most important binding motif determinant.

Most of the free energy contribution, ΔG, of the interaction between positively charged residues on proteins and sulfates on GAGs comes from the entropically favourable displacement of Na^+ or equivalent counter ions bound to the GAG [188-190]. This is called the polyelectrolyte effect. The counter ions have high translational entropy and when they are dispersed into the solvent after complexation, the entropy gain can balance the loss of conformational entropy of the GAG.

Interestingly, arginines and lysines form intermolecular bonds with sulfates in different manners [191], [192] and it has been shown that an arginine peptide binds 2.5 times tighter to heparin than the equivalent lysine peptide [193]. "Stronger hydrogen bonding and a more exothermic electrostatic interaction" likely explained the greater affinity of arginine for GAG sulfates and it is argued that "this can be rationalized by soft acid, soft base concepts" [193].

The type of sugar units the GAG is built up of and their conformation will determine the orientation of the sulfate, carboxyl and hydroxyl groups [174]. These groups are the main interaction points between the GAG and the protein. The interconversion between the IdoA $^1C_4$ and $^2S_0$ conformations only gives rise to small changes in the glycosidic linkage torsions and thus preserves the helical shape of the GAG. On the other hand, the orientation of the sulfate groups in the IdoA residues changes more drastically between the $^1C_4$ and $^2S_0$ conformations (Figure 1.21). The helical shape and the position of the GlcN sulfates remain roughly the same.

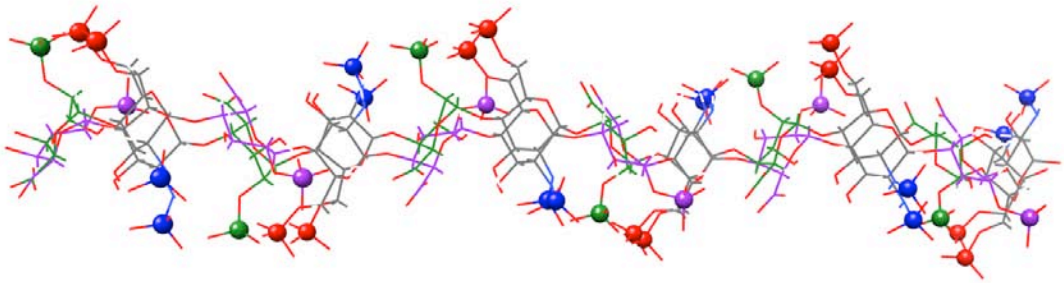
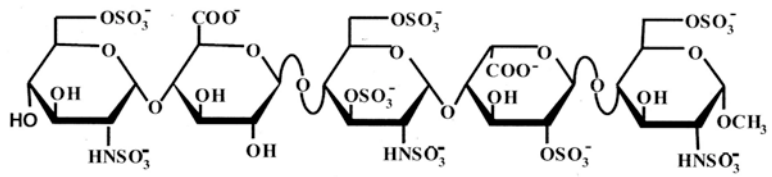


Figure 1.21 This is an overlay of heparin structures, showing the difference of having the IdoA residues in $^{1}C_{4}$ (purple) and $^{2}S_{0}$ conformations (green). Glucosamine residues are shown in grey with the C-2-N-sulfates in blue and the C-6-O-sulfates in red. The figure was created by rendering the 1HPN pdb file [173] in UCSF Chimera 1.4.

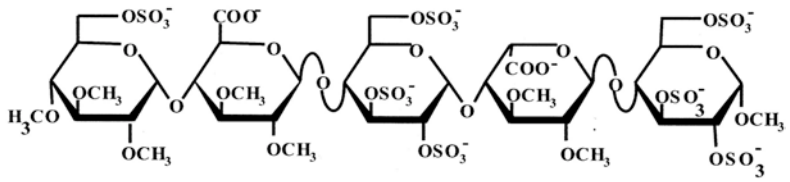
Both the glycosidic torsion angles and sugar conformations often change upon protein binding, creating a distortion in the secondary structure to maximise ionic and van der Waals interactions with the protein [125]. For example, in the FGF-GAG interaction, an iduronic acid residue surrounded by two glucosamines adopts conformations so that the oligosaccharide is kinked to optimise the number of non-covalent bonds to the protein [194]. The IdoA residue is vital in this case as it allows the binding motif to adapt to the protein binding site. The high flexibility of GAGs and their ability to adapt to protein binding sites might explain why some protein-GAG interactions are relatively promiscuous.

It has been shown that heparin interacts specifically with Cu^{2+} [195], Ca^{2+} [196] and many other cations in structurally distinct ways [197]. Interestingly, Ca^{2+} was found to bind specifically between IdoA and GlcN residues where it induced a change in the equilibrium of IdoA conformers towards $^{1}C_{4}$, whereas Na^{+}, K^{+} and Mg^{2+} binding modes were less specific. Further, a heparin oligosaccharide that was unable to support FGF signalling alone strongly supported signalling in the presence of Cu^{2+} [197]. The results emphasise that different cations can promote certain GAG conformations and that this is likely to affect protein – GAG interactions, something that has to be considered when designing experiments.

However, it should also be emphasised that hydrogen bonding, van der Waals interactions and hydrophobic interactions can also be important for some interactions. Fondaparinux and Idraparinux are 3-O sulfate containing HS mimetic anti-coagulant drugs that bind tightly to ATIII (Figure 1.22). It has been shown that the GAG mimetic Idraparinux binds with hundredfold higher affinity to antithrombin than Fondaparinux and that 70% of the affinity of the interaction is made up of non-ionic interactions [198]. This is in strong contrast to the normal antithrombin – heparin interaction, where almost all of the binding affinity can be addressed to ionic interactions [188]. It has been shown that by extending the Fondaparinux oligosaccharide, more ionic interactions can be made, making the ionic contribution even greater [189]. By performing experiments at different temperatures, in buffers with different heats of ionisation and salt concentrations, it is possible to quantify the ionic bond, hydrogen bond and hydrophobic interaction contributions. It is also possible to quantify the number of hydrogen and ionic bonds that are formed in the binding event, how many sodium ions that are displaced and the net uptake or release of hydrogens from the buffer.



FONDAPARINUX



IDRAPARINUX

Figure 1.22 Structures of Fondaparinux and Idraparinux. Taken from [199].

1.3.8 Enzymatic Depolymerisation of Glycosaminoglycans

GAG chains are typically 50-100 monosaccharide units long and commercial heparin varies in size from 5-40 kDa, although size and other parameters vary between batches [200], [126], [201], [69]. Long GAGs often precipitate heparin binding proteins, it can therefore be necessary to digest them into shorter fragments to be able to study the interaction. Further, the heterogeneity can be reduced and studied to decipher the disaccharide sequence code [202]. There are several different ways to depolymerise GAGs, the most gentle being enzymatic digestion, but chemical depolymerisation is also commonly employed to produce GAG fragments [203].

GAG degrading enzymes exist in both eukaryotes and prokaryotes, although many of the enzymes used in research and for production of pharmaceutical products come from the gram-negative soil bacterium *Flavobacterium heparinum*, also called *Pedobacter heparinus* [204]. Bacteria can use GAGs not only as a carbon source, but also as a sulphur source [205].

Three different heparinases (I-III) with different specificities have been cloned, overexpressed and purified from *Pedobacter heparinus* [206]. They all digest heparin and HS by cleaving the glycosidic linkage between a glucosamine and an uronic acid through a β-elimination reaction, producing a double bond on the non-reducing end of the oligosaccharide [207] (Figure 1.23).

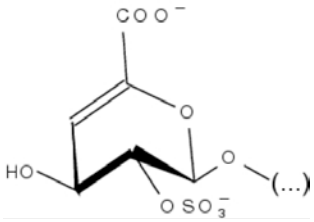


Figure 1.23 The non-reducing end sugar unit produced by heparinase I and II.

Conveniently, the non-reducing end double bond is conjugated with the carboxylic acid group and therefore absorbs UV light at 232nm, which can be used for quantitation [208]. The catalytic mechanism of heparinase is calcium dependent and calcium has to be included in the digestion buffer [209]. Heparinase I cleaves heparin and HS between $GlcNSO_3$ and IdoA-$2SO_3$, heparinase II is more promiscuous and cleaves between glucosamine and both kinds of uronic acid residues and heparinase III which is selective for lowly sulfated regions of HS cleaves between $GlcNSO_3$ or GlcNAc and glucuronic acid residues (Figure 1.24) [210]. There are also chondroitinases to digest CS and DS (Figure 1.24) [211-213]. Chondroitinases are not calcium dependent.

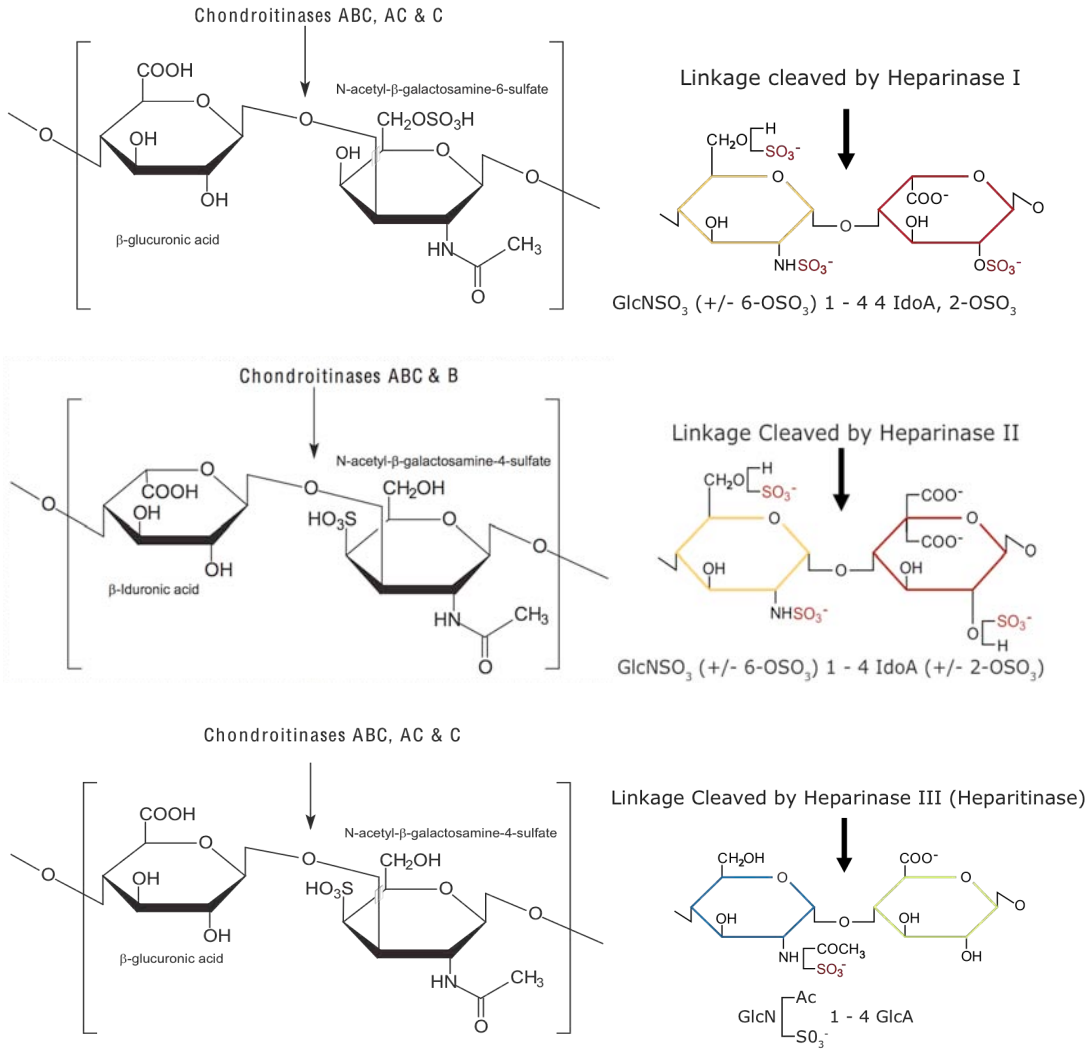


Figure 1.24 Specificities of the different chondroitinases are shown on the left. From the top, the GAGs are CSA, DS and CSC. Chondroitinase B refers to the old name

for dermatan sulfate, chondroitin sulfate B. The different heparin disaccharides that heparinases digest are shown on the right. Taken from [214] and modified from [215].

The enzyme digest can be stopped before it goes to completion and the produced oligosaccharide sizes can be separated by gel filtration. After gel filtration, strong anion exchange chromatography (SAX) can be employed to isolate fully sulfated species from those that lack sulfates. The use of SAX chromatography is only practically realistic for shorter heparin/HS oligosaccharides. Individual heparin/HS tetrasaccharides can be readily purified and some individual hexasaccharide species can be separated, but for octa to deca –saccharides, only fully sulfated species can be purified. If homogeneous, fully sulfated oligosaccharides are purified, most of the material will be lost due to the heterogeneity.

The GAG lyases are also important in GAG sequencing techniques and disaccharide analysis [168], [169]. Disaccharide analysis plays a central role in unlocking the structural basis of protein-GAG interactions and identifying the binding motif. The technique relies on digesting the GAG oligosaccharide of interest into its disaccharide components and comparing the disaccharides' elution times on capillary electrophoresis, SAX or reverse phase HPLC to known standards. By comparing the sample disaccharides to the standard disaccharides, the disaccharide composition will be clarified, but not necessarily the sequence in which the disaccharides are linked. To elucidate the sequence, incomplete digests or NMR techniques have to be employed. Thus, the technique is particularly powerful when coupled to other analytical techniques such as NMR and mass spectrometric techniques [216].

1.3.9 Glycosaminoglycan Research

More and more is learnt about the importance of GAGs in many physiological functions. Figure 1.25 shows how the number of GAG related publications has grown linearly during the last 28 years. PubMed publications have been growing

faster compared to the glycosaminoglycan related publications in the last 20 years and show signs of a more exponential like growth curve, showing that other research topics have attracted more interest.

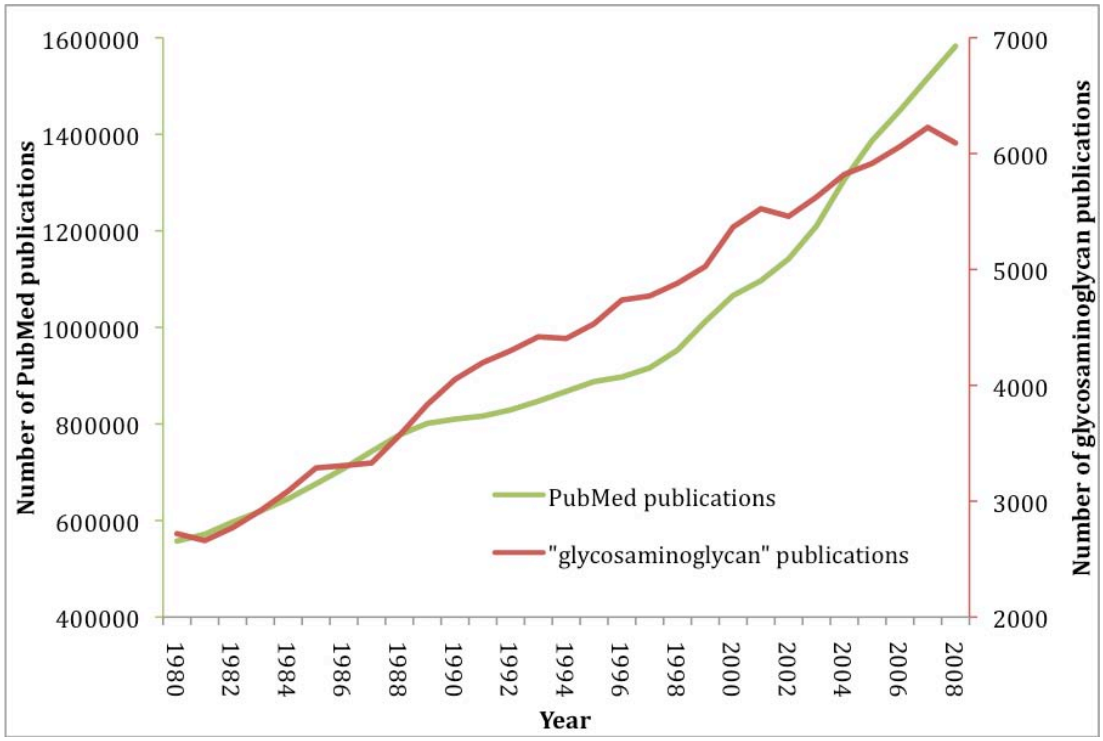


Figure 1.25 The graph shows number of total PubMed publications per year on the left axis and number of glycosaminoglycan publications per year on the right axis sorted per year.

2 Aims

Hepatocyte Growth Factor is a multifunctional growth factor that controls many physiological processes, for example, cell proliferation, growth and morphology. HGF exerts its biological functions by binding Met, a tyrosine kinase receptor that activates intracellular signalling cascades. There is extensive evidence of HGF's involvement in disease, especially in cancer and in benign physiological processes such as wound healing. It has been suggested that glycosaminoglycans such as heparan sulfate (HS) and dermatan sulfate (DS), help or are required in the formation of a ternary active GAG-HGF-Met signalling complex. The HGF-GAG interaction is therefore a target for drug development, in particular for anti-cancer therapy. One potential approach is to use GAG-mimetics that bind HGF, but do not activate Met. Such drugs are currently being developed by pharmaceutical companies. In order to successfully target protein-GAG interactions for drug development, an in-depth understanding of these interactions is required. The overall aim of this thesis was therefore to contribute to the molecular-level understanding of the HGF-GAG interaction.

In order to achieve the aim of clarifying which protein residues that are involved in the interaction and how these interact with different GAGs, recombinant proteins and enzymatically depolymerised GAGs had to be produced. The goal was to study the interaction from both a protein and GAG perspective, establishing which structural features of the GAGs and HGF that are important for this interaction.

Although HGF-NK-G has been crystallised in the presence of heparin DP14, all aspects of the HGF-heparin interaction are still not understood. For example, the protein construct used in the crystal structure contains an N-terminal cloning artefact, which introduces negatively charged residues near the GAG binding site. It was therefore judged important to use a more native protein to study the interaction and determine if this cloning artefact affects the binding mode of the GAG. Heparin and DS are the only GAGs that have previously been used in molecular-level studies of

HGF-GAG interactions. Therefore, another important aim of this thesis was to elucidate any differences in binding mode of an extended set of GAGs under identical conditions using solution-based techniques.

Ionic intermolecular bonding is typically the most important force in protein-GAG interactions and therefore, many proteins have a preference to bind to NS region over NA regions in heparin/HS. It has been suggested that this is also the case for HGF. However, limited data is available to answer the question whether a specific motif or sulfation patter in a heparin/HS oligosaccharide binds tighter than others. A goal of this thesis was therefore to investigate which heparin sequence in a mixture of heparin oligosaccharides that binds tightest to HGF-N.

The K_Ds for the interaction between GAGs, HGF and its NKx splice variants have been determined mostly by biosensor/SPR analysis. Several experimental designs with different immobilisation techniques have previously been employed in these studies. The biosensor/SPR results suggest that the K_D of the interaction is in a low nM range. Some protein-GAG interactions are of high affinity with K_Ds in a low nM range, for example the ATIII-pentasaccharide and bFGF-heparin interactions, whereas most other protein-GAG and protein-carbohydrate interactions have K_Ds in the μM range. Solution based methods and preliminary ITC studies of the HGF – heparin interaction have suggested that the interaction in a low μM range. A key aim of this thesis was therefore to investigate the affinity of the interaction further with solution based techniques, in particular NMR and ITC to contribute to the understanding of the affinity of the interaction and whether the K_D of the interaction is in a low μM range or low nM range.

It has been proposed that HGF and in particular the NK1 protein construct oligomerises upon GAG binding and that the oligomerisation process can be involved in the action mechanism of HGF. The oligomerisation process has previously not been studied in great detail and one aim of this thesis was to extend

the oligomerisation studies to different types and lengths of GAGs and to investigate the oligomerisation mode of HGF. This includes the identification of the shortest GAG oligosaccharide that is sufficient to induce oligomerisation and how oligosaccharide length and type affects oligomerisation mode.

It has been demonstrated in the past that linear carbohydrates, as asymmetric molecules, can bind to proteins in different orientations (reducing-to-non-reducing end or non-reducing-to-reducing end). One of the aims of this thesis was therefore to clarify if GAGs bind to HGF in a preferred direction. To achieve this aim, a spin-labelled heparin oligosaccharide was prepared and titrated into ^{15}N labelled HGF. Monitoring of proton relaxation times of backbone NH resonances allowed this question to be answered.

3 Materials and Methods

3.1 Analytical Techniques

3.1.1 Quantification of Proteins and GAGs

Protein extinction coefficients were calculated by using ProtParam (http://expasy.org/tools/protparam.html) and the protein concentrations were determined by A_{280}. The extinction coefficients for the protein constructs used in this thesis are given in Table 3.1.

Table 3.1 Molar extinction coefficients for the different protein constructs.

Construct	**Molar extinction coefficient**
HGF-N	8730 M^{-1} cm^{-1}
HGF-K	17335 M^{-1} cm^{-1}
HGF-NK-C	26065 M^{-1} cm^{-1}
HGF-NK-G	27555 M^{-1} cm^{-1}

All enzymatically digested GAG oligosaccharides were quantified by A_{232}, using a molar extinction coefficient of 5200 M^{-1} cm^{-1} [208].

3.1.2 SDS-PAGE

Protein samples were mixed with Invitrogen NuPage 4X loading buffer (NP0007), 10X reducing agent (NP0004) and heated to 75 ºC for 10 minutes. The final samples were run on a 4-12 % NuPage Bis-Tris gel (NP0321BOX) with MES (NP0002) or MOPS (NP0001) running buffer at 250 volt for 40 minutes and then stained with Bio-Rad BioSafe Commassie Stain (1610787).

3.1.3 Akta FPLC System

The Akta FPLC system (GE Healthcare) consisted of a UPC900 UV unit monitoring the 280 nm trace, a P920 pump unit and a Frac-900 fraction collector, all controlled via the UNICORN 3.21 software.

3.1.4 BioCAD FPLC System

The FPLC instrument was a BioCAD 700E system with a Waters fraction collector capable of collecting 120 fractions.

3.1.5 Waters HPLC System

The Waters HPLC system consisted of a 486 tuneable UV detector, a 600 control unit and a 616 pump unit controlled via the Millennium 3.05.01 software and a 120 fractions Waters fraction collector that was programmed manually.

3.1.6 NMR Spectrometers

The AVA600 spectrometer was a Bruker Avance 14.1 T spectrometer (600 MHz proton Larmor frequency) equipped with a 5-mm triple resonance TXI probe.

The BIO600 spectrometer was a Bruker Avance 14.1 T spectrometer (600 MHz proton Larmor frequency) equipped with a 5-mm triple resonance TXI cryoprobe.

The AVA800 spectrometer was a Bruker Avance 18.8 T spectrometer (800 MHz proton Larmor frequency) equipped with a 5-mm triple resonance TCI cryoprobe.

3.1.7 Processing of NMR Data

NMR data was acquired using Bruker TOPSPIN 1.2 patch 8. Spectra were processed using the process program from the AZARA 2.7 suite and then imported into CcpNmr Analysis 1.0 R15 [217] for assignment and data analysis. The assignment of

^{15}N-^{1}H HSQC crosspeaks of HGF-N was achieved by comparison of the spectra with a previously published assignment [17].

3.1.8 MALDI-TOF Mass Spectrometry of Proteins

Spectra were collected on an Applied Biosystems DE-STR MALDI-TOF with a nitrogen laser. Sinapinic acid was used as matrix for protein samples and prepared by sonicating and vortexing 10 mg sinapinic acid from Sigma-Aldrich (85429-1G) in 1 ml 50 % acetonitrile/ddH_2O mixture with 0.03 % TFA. 0.5 μl protein sample was mixed with 0.5 μl matrix directly on the target plate. The lowest laser intensity that gave a clean peak was used. When needed, the instrument was calibrated with the external standards cytochrome c, horse myoglobin and hen egg white lysozyme.

3.1.9 MALDI-TOF Mass Spectrometry of GAGs

To collect MALDI-TOF spectra of GAGs, 25 mg/ml 3-Hydroxypicolinic acid from Sigma-Aldrich (56197-250MG) was dissolved in 50% Acetonitrile/ddH_2O. No acid was added to the matrix solution. A RGRGRGRGRGRGRGRGRGRGRG ($(RG)_{11}$) peptide was synthesised by Peptide Protein Research Ltd (http://www.peptidesynthetics.co.uk/). A small volume of fresh matrix with 5-50 μM peptide (typically 10 μM) was prepared fresh before each experiment. It was found that the $(RG)_{11}$:GAG ratio was very important to give high quality spectra and typically one to ten times excess of $(RG)_{11}$ peptide was used. Occasionally, this was a problem when only a small sample volume with an unknown GAG concentration was available. To get the right $(RG)_{11}$:GAG ratio, ten matrix-$(RG)_{11}$ solution spots were pipetted onto the target plate, the sample was mixed with the first spot and nine 1:1 dilutions were made on the remaining spots. High quality spectra could be collected when the matrix crystallised as long rods, if the matrix crystallised as dots, spectra could usually not be collected at all. The method of using peptide-GAG complexes for MALDI-TOF analysis of GAGs has been described before [218-221].

Spectra were collected on the same instrument as for protein MALDI-TOF in positive reflectron mode with 65% grid voltage, 25 000 accelerating voltage, 225 ns

extraction delay time and 0.05% guide wire. Higher laser intensity had to be used compared to proteins and peptides. For proteins, a laser intensity of approximately 1700 was used, but for GAGs a laser intensity in the range of 2100 gave good spectra.

3.2 Protein Production

3.2.1 Cloning and Transformation of HGF-N, HGF-K and HGF-NK-C into P. pastoris

Image Consortium cDNA clone 40146317, corresponding to Entrez nucleotide bc130286.1 was obtained from Geneservice (Cambridge, UK) in a PCR4-TOPO vector. A Stratagene Quickchange kit was used to make the silent mutation t615a to remove a XHOI restriction site as XHOI was going to be used to insert the gene fragment into the pPICZalphaB expression vector.

Primer set for mutagenesis was:

5' (Forward) Primer: 5'-CTACTGTCGAAATCCACGAGGGGAAGAAGGG-3'

3' (Reverse) Primer: 5'-CCCTTCTTCCCCTCGTGGATTTCGACAGTAG-3'

The Stratagene Quickchange mutagenesis was performed as described by the manufacturer and the DPNI digested PCR product was transformed into chemically competent TOP10 *E. coli* cells. 20 μl and 200 μl aliquots of the transformation reaction were plated and incubated over night on 100 μg/ml LB-ampicillin plates. The day after, 10 colonies were inoculated in 5 ml LB broth containing 100 μg/ml ampicillin. Next, plasmid DNA from the 10 colonies was prepared by using the Qiagen QIAprep Spin Miniprep Kit following the manufacturer's protocol, which typically yielded 50 μl 250 μg/ml DNA. To screen for clones containing the gene insert, 10 μl DNA was incubated with 0.5 μl (10 units) XHOI restriction enzyme from New England Biolabs (R0146S) and run on 1% agarose gel stained with SYBR Safe from Invitrogen (S33102). The DNA for two clones that ran higher than uncut control DNA was sequenced using M13 primers and BigDye Terminator v3.1 from

Applied Biosystems. The sequence verified the T615A mutation and the clone was inoculated in a 5 ml LB-ampicillin Falcon tube overnight. The following morning, 850 μl of the culture was mixed with 150 μl autoclaved glycerol and snap-frozen in liquid nitrogen to produce a glycerol stock stored at -80 °C.

Next, the different HGF fragments were amplified by PCR and later cloned into the pPICZalphaB expression vector. Forward primers contained 4 random bases followed by an XHOI restriction site, then the KEX2 cleavage site glutamic acid, arginine and lysine residues followed by the protein sequence. Reverse primers contained the last codons of the cloned sequence, a TAA stop codon followed by an XBAI restriction site and four random residues. The four random bases improve restriction enzyme activity. The KEX2 arginine and lysine had to be rebuilt with PCR when using the pPICZAlphaB XHOI restriction site (Figure 3.1). Refer to Table 3.2 for a complete list of the different primers and protein constructs.

The forward primers for HGF-N and HGF-NK-C (Table 4.1) were chosen so that the first residue of the expressed protein was 31G. The first residue of native HGF is 32Q, but 31G was chosen as the first residue because that would produce the same protein as used for the determination of the NMR structure by Zhou *et al* [21], making it uncomplicated to copy the assignment of ^{15}N-^{1}H crosspeaks.

The α-factor signal sequence was used to direct the expressed protein for secretion. The Kex2 enzyme in *P. pastoris* cleaves the α-factor signal sequence as the expressed protein is secreted between Arg and Glu in the sequence Glu-Lys-Arg * Glu-Ala-Glu-Ala, where * is the cleavage site. The Glu-Ala repeats are not necessary for Kex2 cleavage, but often makes the cleavage more efficient. The Glu-Ala repeats are further cleaved by the Ste13 gene product. For some expressed proteins, Ste13 cleavage of Glu-Ala repeats is not efficient and Glu-Ala repeats are left on the N-terminus. For this reason, the protein constructs in this thesis were cloned without the Glu-Ala repeats.

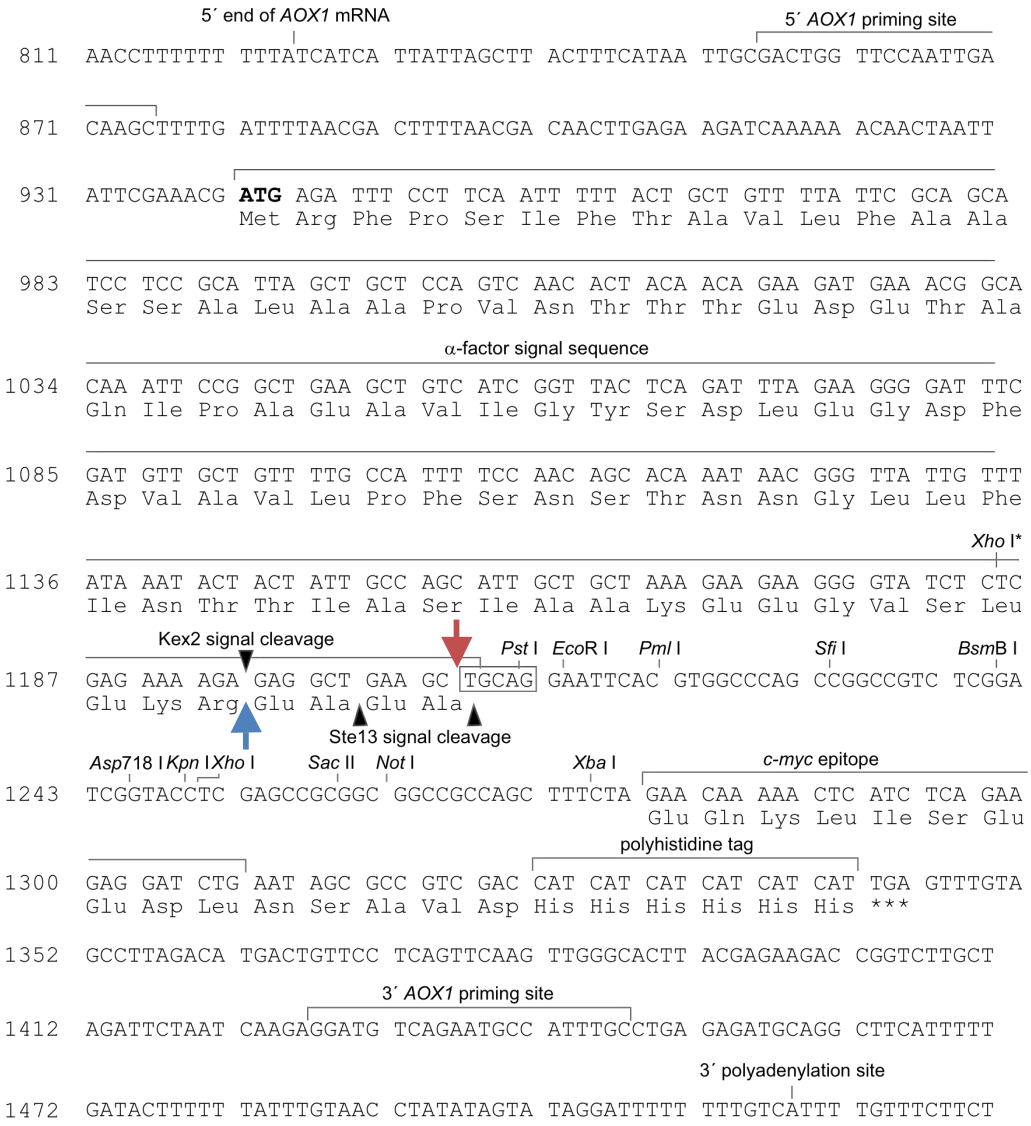


Figure 3.1 The pPICZAlphaB multiple cloning site. The blue arrow indicates where the protein sequence was inserted in this project. The red arrow indicates where the protein sequence would have been inserted if the Glu-Ala-Glu-Ala cloning artefact would have been included in the protein. The artefact is removed post-translation by the Ste13 enzyme, but cleavage is sometimes inefficient, leaving one or two Glu-Ala repeats on the N-terminus.

Table 3.2 Protein constructs cloned and expressed in this thesis.

Construct	Primer Set	Residues	Protein Sequence
HGF-N	**5' (Forward) Primer**	31-127	GQRKRRNTIH EFKKSAKTTL IKIDPALKIK
	5'-GAACTCTCTCGAGAAAAGAGGACAAAGGAAAAGAAG-3'		TKKVNTADQC ANRCTRNKGL
	3' (Reverse) Primer		PFTCKAFVFD KARKQCLWFP
	5'-GACCTCTAGATTAGTTTCTAATGTAGTCTTTG-3'		FNSMSSGVKK EFGHEFDLYE NKDYIRN
HGF-K	**5' (Forward) Primer**	126-208	RNCII GKGRSYKGTV SITKSGIKCQ
	5'-GAACTCTCTCGAGAAAAGAAGAAACTGCATCATTG-3'		PWSSMIPHEH SFLPSSYRGK
	3' (Reverse) Primer		DLQENYCRNP RGEEGGPWCF
	5'-TCATTCTAGATTATTCTGAACACTGAGGAATG-3'		TSNPEVRYEV CDIPQCSE
HGF-NK-C	**5' (Forward) Primer**	31-208	GQRKRRNTIH EFKKSAKTTL IKIDPALKIK
	5'-GAACTCTCTCGAGAAAAGAGGACAAAGGAAAAGAAG-3'		TKKVNTADQC ANRCTRNKGL
	3' (Reverse) Primer		PFTCKAFVFD KARKQCLWFP
	5'-TCATTCTAGATTATTCTGAACACTGAGGAATG-3'		FNSMSSGVKK EFGHEFDLYE
			NKDYIRNCII GKGRSYKGTV SITKSGIKCQ
			PWSSMIPHEH SFLPSSYRGK
			DLQENYCRNP RGEEGGPWCF
			TSNPEVRYEV CDIPQCSE

A 50 μl PCR reaction containing 5 ng template HGF DNA, 1 nM (1 μl 100 μM) of each primer, 1 μl 10 μM Bio-Rad dNTP mix and 1 μl Herculase DNA Polymerase (Stratagene) was set up with the cycling conditions in Table 3.3.

Table 3.3 PCR cycling conditions for amplifying DNA to be cloned into expression vector.

Temperature	**Time**
95 °C	1:30
95 °C	30
60 °C	30
72 °C	45
Goto 2 Rep	35
72 °C	8:00
4 °C	Hold

Amplification was confirmed by running an aliquot of the DNA on an agarose gel. The template vector was cut with SPEI and AlwNI to make sure that it wouldn't contaminate the cloning reaction and giving false positives. Ideally the template and target vector should have carried different antibiotic resistances as some false positives were still detected. A 2 μl aliquot of the digested PCR reaction was used to clone the PCR product into a TOPO vector using a Zero Blunt TOPO PCR Cloning Kit from Invitrogen, following the protocol. After that, 3 μl of the TOPO cloning reaction was transformed into TOP10 cells. To screen for positive TOPO cloning reactions, 8 colonies of each construct were diluted in 200 μl ddH_2O and 9 μl of this cell suspension was used as DNA template in a PCR reaction with 1 μl gene specific

primer and 10 μl PCR Master Mix from Promega with the cycling conditions in Table 3.4.

Table 3.4 PCR cycling conditions for screening *E. coli* clones containing the correct inserts.

Temperature	**Time**
95 °C	8:00 (to lyse cells)
95 °C	30
58 °C	30
72 °C	60
Goto 2 Rep	35
72 °C	8:00
4 °C	Hold

The PCR reactions were loaded onto an agarose gel to confirm that the inserted DNA was of the correct length. Five clones of each construct were grown overnight, minipreped and sequenced. One positive clone of each construct was used to make a glycerol stock.

Next, each TOPO construct was prepared for cloning into the *Pichia* expression vector pPICZAlphaB from Invitrogen by digesting 21.5 μl DNA (~250 ng/ml) with XHOI and XBAI. The reaction was stopped by heating to 65 °C for 20 min.

An *E. coli* clone carrying empty pPICZAlphaB vector was grown over night in 250 ml of LB broth with 50 μg/ml Zeocin. The culture was used to maxiprep DNA with a Sigma-Aldrich GenElute HP Kit. The prepared DNA was digested with XHOI and

XBAI. After that, 50 µl pPICZAlphaB and gene insert DNA was gel extracted with a Qiagen Qiaquick Gel Extraction Kit. To further purify the gel extracted DNA, it was phenol:chloroform extracted with Phenol:Chloroform:Isoamyl Alcohol 25:24:1, 10 mM Tris, pH 8.0 from Sigma-Aldrich. To concentrate and purify the DNA further, it was ethanol precipitated in 100% EtOH with 0.1 volume 3 M Sodium Acetate pH 5.2 and washed with 70% EtOH.

The pPICZAlphaB vector and the cDNA genes were mixed at a 1:4 ratio and ligated using a Quick Ligation Kit from NEB, following the protocol. The ligated DNA was transformed into *E. coli* TOP10 cells and screened for correct insertion by PCR, amplifying the inserted cDNA with PCR MasterMix and primers that bound 5' and 3' of the gene insert in pPICZAlphaB (Table 3.5). Positive clones were submitted for sequencing and glycerol stocks were made of positive clones.

Table 3.5 pPICZAlpha sequencing primers.

5' pPICZAlpha sequencing primer	GGGGATTTCGATGTTGCTGTT
3' pPICZAlpha sequencing primer	CCGGTCTTCTCGTAAGTGCC

To transform the finalised HGF-N, HGF-K and HGF-NK-C pPICZAlphaB expression vectors into *P. pastoris*, the three vectors were maxipreped and 100 µg DNA was linearised with SACI from NEB (R0156S) in a volume up to 500 µl. A 1 µl aliquot of the digested DNA was analysed on a 0.5% agarose gel to confirm that the DNA was properly linearised. The DNA was then phenol:chloroform extracted 3 times, ethanol precipitated and dissolved in 22 µl ddH_2O. A 2 µl aliquot of the linearised and digested DNA was used to assess concentration and purity of the DNA.

Electrocompetent KM71H *P. pastoris* cells were prepared according to the manual [222], with the exception that the first two centrifugation steps were performed at 1000 g instead of 1500 g to give the cells a softer treatment and make the pellet easier to resuspend. Usually 20, 40 and 80 µl starter culture in late exponential phase were used to inoculate 100-500 ml YPD the night before. Three different inoculation volumes were used to make sure that the cells could be prepared at the right OD at the right time. For transformation, 10 µl DNA (roughly 10 µg) was mixed with 80 µl cell suspension, incubated on ice for circa 3 minutes and electroporated with a BioRad Gene Pulser II according to the manufacturers recommendations (0.2 cm cuvette, 1.5 kV voltage, 7.5 kV/cm field strength, 25 µF capacitor, 400Ω resistor, 8 msec time constant). The cells were resuspended in 1 ml 1 M sorbitol and incubated for two hours at 30 ºC. To select transformants and screen for multi-copy mutants at the same time, 250 µl of the cell suspension was plated on 100 - 300 µg/ml YPDSZ plates.

Molecular cloning of all constructs into the pPICZAlphaB expression vector was successful. It is a tedious procedure and recent advances in codon optimisation and DNA synthesis make full codon optimised gene synthesis a more attractive approach [223]. PCR screening was employed to identify successful cloning of the protein sequences into the expression vector for all three protein constructs (Figure 3.2). Primers that bind 5' and 3' of the gene insert in the pPICZAlphaB vector were used to amplify DNA. High running bands confirm the presence of inserted DNA, low running bands are from clones not containing the correct DNA.

Figure 3.2 Colony PCR screening for clones with correctly ligated gene inserts into pPICZAlphaB. Each lane is one clone. The first two lanes are control PCR reactions. An empty lane separates the HGF-N and HGF-K constructs. Low running bands have migrated through the gel relatively quickly, which indicates a smaller size and therefore, they do not contain the inserted DNA. High running bands have migrated slowly through the gel slowly due to their relatively large size, which indicates that they contain the inserted DNA.

DNA sequencing was used to confirm that all protein sequences were correctly cloned into the pPICZAlphaB vectors (Figure 3.3).

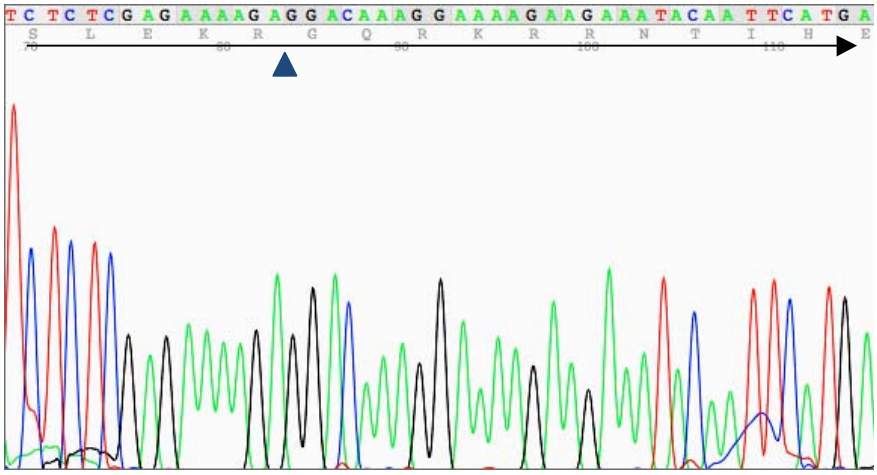


Figure 3.3 Sequencing chromatogram showing HGF-N correctly cloned into pPICZAlphaB. The DNA sequence can be read on the top, and the translated sequence is in grey. The KEX2 cleavage site is indicated with a blue arrow. The first residues of the secreted protein are GQKR just after the KEX2 cleavage site.

pPICZAlphaB vectors with protein inserts were linearised with SACI before being transformed into *P. pastoris*. An agarose gel was run to confirm that the digest had worked and that the vector was completely linearised (Figure 3.4). Supercoiled vector DNA migrates much faster than linear digested vector DNA through an agarose gel and will therefore have a lower apparent kilobase. This principle was used to compare digested and undigested DNA to judge if the vector was successfully linearised.

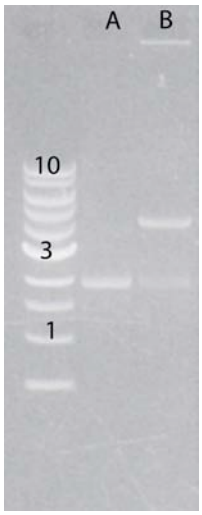


Figure 3.4 A 0.7 % agarose DNA gel of HGF-N cloned into pPICZAlphaB for transformation. The first lane is a kilobase ladder and the numbers indicate the number of kilobases in the ladder. The second lane (A) is undigested control vector and the last lane (B) is linearised vector, digested with SACI.

3.2.2 Selection of Clones, Expression and Purification of HGF-N, HGF-K and HGF-NK-C

To identify potential multi-copy insert clones that express protein, several clones of each construct were screened for protein expression on a small scale. After transformation, the *Pichia* cells were incubated for 4 days at 30 ºC and 10 of the largest colonies on the highest Zeocin concentration plates were inoculated in 10 ml BMG Falcon tubes and incubated at 30 ºC, 250 rpm for 2 days. The tubes were then centrifuged for 5 minutes at 1000 g and resuspended in 1-2 ml BMM. The cultures

were fed 1 % Methanol every 24 hours for 4 days. To harvest the cultures, tubes were centrifuged at 4000 g for 20 min and 750 µl of the supernatant was concentrated 10-15 times in a Sartorius Vivaspin 500 5000 MWCO concentrator (FIL8550). Next, 15µl of the concentrated supernatant was analysed by SDS-PAGE. The clones that showed the strongest protein band of the right size after SDS-PAGE analysis were inoculated in 5 ml YPD broth to make glycerol stocks.

After identifying the best expressing clone for each protein construct, the proteins were expressed in a fermenter to yield larger protein quantities. A 10 ml BMG starter culture was incubated for one day and used to inoculate a 200 ml BMG shaker that was incubated for one more day. The shaker culture was centrifuged for 5 minutes at 1500 g and the pellet was resuspended in 30 ml sterile PBS and used to inoculate a fermenter with fermentation minimal medium (Table 3.6 and Table 3.7).

Table 3.6 Unlabelled fermentation minimal medium.

Phosphoric acid, 85 %	27 ml/l
$CaSO_4$	0.95 g/l
K_2SO_4	18.2 g/l
$MgSO_4$ x $7H_2O$	15 g/l
KOH	4.2 g/l
Glycerol	25 ml/l
dH_2O	Up to 1 litre
After autoclaving, 4.35 ml PTM1 salts and 0.5 ml AntiFoam 204 (Sigma-Aldrich) per litre medium were added.	

Table 3.7 ^{15}N-labeled fermentation minimal medium.

$CaSO_4$	0.75 g
$MgSO_4$ x $7H_2O$	12 g
K_2SO_4	8 g
1M KPO_4 pH 6.0	60 ml
Glycerol	20 ml
dH_2O	Up to 600 ml
After autoclaving, 4.35 ml PTM1 salts and 0.5 ml AntiFoam 204 (Sigma-Aldrich) per litre medium were added.	

A New Brunswick Bioflow3000 system equipped with a pH probe and a dissolved oxygen probe from Mettler Toledo was used for fermentation. A Windows PC running the BioCommand application controlled, monitored and recorded the fermentation. For unlabelled fermentations, a 10-litre vessel was used, for ^{15}N-labelled fermentations a 2-litre vessel was used. The dissolved oxygen set point was 40 %, minimum agitation 200 rpm, maximum agitation 1000 rpm and temperature was maintained at 30 °C for growth. A control loop was setup so that dissolved oxygen was controlled firstly by agitation increasing from 200 to 1000 rpm followed by enriching the airflow with increasing concentrations of O_2. The pH was maintained at pH 5 by addition of 0.88 M NH_4OH for unlabelled fermentations and 2 M KOH for ^{15}N-fermentations. For unlabelled fermentations an initial volume of 2.5 lit was used and 250 ml 50% glycerol with 1.5 ml PTM1 salts was added three or four times when the cells had consumed all the glycerol. For ^{15}N labelled fermentations, the initial medium volume was 600 ml, 5 g $(^{15}NH_4)_2SO_4$ in 50 ml dH_2O and 40 ml 50% glycerol with 0.5 ml PTM1 salts was added when the medium was inoculated.

Before induction of protein expression, the cells were allowed to consume all glycerol and the temperature was set to 15 °C. The first methanol feed was 0.5% and

the cells slowly consumed the feed as they started to produce AOX enzyme. Subsequent methanol feeds were 1.5% and the cells consumed the methanol much quicker as the AOX enzyme was already present. It typically took 3-6 hours for the cells to consume each methanol feed, as indicated by an increase in dissolved oxygen and a decrease in agitation. The methanol contained 4.4 ml PTM1 salts/L methanol.

After 4 days of methanol feeding, the culture was harvested by centrifugation at 5,000 rpm for 15 minutes. The supernatant was centrifuged again at 10 000 g for 30 min and filtered through a 0.45 μm PES filter. The filtrate was diluted 1 to 4 with dH_2O and 5 mM EDTA, 0.5 mM PMSF and 0.02% w/v sodium azide were added to inhibit proteolysis and bacterial growth.

High Zeocin resistant clones were successfully screened for protein expression for all protein constructs. After expression had been confirmed by miniscale SDS-PAGE analysis, the clones were fermented in a bench-top fermenter to yield larger quantities of protein.

In the fermenter, agitation rises more and more after inoculation, as the cells grow and consume oxygen (Figure 3.5). After the cells had consumed all the initial glycerol, additional glycerol was fed in twice as can be seen by a rise in agitation at 30 and 50 hours. The green line in Figure 3.5 shows how the cells produced acid as they grew on glycerol, but not on methanol: initially ammonia had to be added to keep the pH at the set point. The temperature was relatively stable; a decrease can be seen as the temperature (cyan line) is set to 15 °C to reduce proteolysis before methanol induction. The temperature increased somewhat as the cell metabolism increased every time the culture was fed methanol. The agitation also increased every time the culture is fed, except in the end when the culture was fed so often that there was always methanol present for the cells to feed on. The DO remained relatively constant: in this fermentation it was lowered half way, as the initial set point was too high.

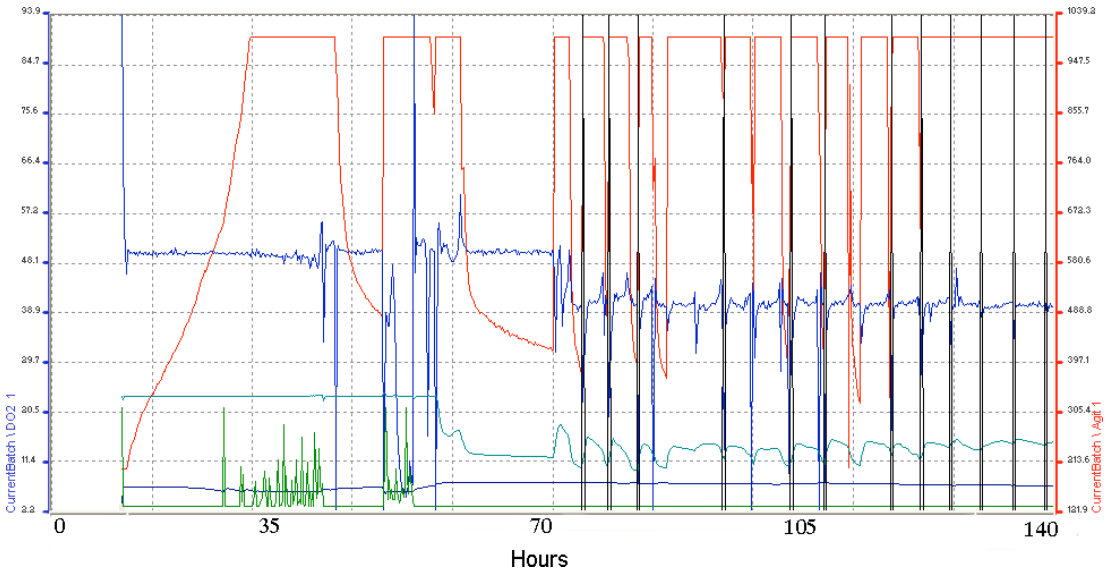


Figure 3.5 Data chart from expressing HGF-NK-C in the fermenter. Red line: agitation, blue line: dissolved oxygen, green line: feed of ammonia, cyan line: temperature, black line: feeding of methanol.

As the *Pichia* cells were fed methanol, the expressed protein was secreted into the medium. After fermentation, the protein was captured on a heparin or SP -Sepharose resin and eluted, collecting fractions and monitoring the A_{280} trace.

To capture, purify and concentrate HGF-N and HGF-NK-C, the filtrate was pumped over night with a peristaltic pump through a 20 cm long Pharmacia XK26 column packed with 12.5 ml Heparin-Sepharose Fast Flow resin from GE Healthcare. HGF-K was captured in the same way on Fast Flow SP-Sepharose resin. For HGF-N and HGF-NK-C, the pH of the filtrate was adjusted to pH 6 and for HGF-K to pH 4.2 to allow selective binding to the resin.

HGF-N and HGF-NK-C were first purified by heparin affinity chromatography and then by strong cation exchange chromatography. For HGF-N and HGF-NK-C, buffer A was 20 mM KPO_4 pH 7.5 with 0.02% (v/v) sodium azide, buffer B was the same, but contained 2 M NaCl.

The loaded XK26 heparin-Sepharose column was washed with buffer A and then eluted with a 5 CV gradient to buffer B on the BioCAD FPLC system at a flow rate of 5 ml/min, monitoring the absorbance at 280 nm, collecting 3 ml fractions. Peak fractions were pooled and subjected to strong cation exchange chromatography using a MonoS 4.6/100PE column (GE Healthcare) connected to the Akta FPLC system with the following gradient: 0-15 % B in 1CV, 15-60 % B in 30 CV, wash with 100 % B and equilibrate.

HGF-K was first purified by SP-Sepharose chromatography in the same manner as HGF-N and HGF-NK-C, but the pH of the buffers was adjusted to pH 6.5. SP-Sepharose fractions were pooled and purified by gel filtration. For gel filtration, a XK16/60 Superdex HiLoad from GE healthcare was connected to the Akta FPLC system described above, with 1 ml/min PBS flow rate, collecting 2 ml fractions.

Protein purity was assessed by SDS-PAGE and MALDI-TOF mass spectrometry, correct amino acid sequence was confirmed by FTICR-MS and correct folding was confirmed by ^{1}H-^{15}N HSQC NMR.

3.2.3 FTICRMS Characterisation of HGF-N and its Modified Versions

200 μl, 20-50 μM samples were dialysed against 2 litres of 50 mM NH_4OAc buffer with 1000 MWCO tube-o-dialyzers from WebScientific (BE786-610) at 4 °C for ~24 hours. Samples were then reduced with an 8X molar excess of TCEP, acidified and 50 % methanol added. The samples were directly infused with a nESI interface for mass determination and ECD in a Bruker Daltonics 12 T FTICR instrument. The mass spectrometry analysis was conducted by Jenna Scotcher, SIRCAMS.

3.3 Preparation and Purification of GAG Derived Oligosaccharides

Many different conditions for GAG digestion and purification were attempted; the conditions presented here are the preferred ones. Typically, 100-1000 mg of bovine lung sodium salt heparin from Calbiochem (375093), bovine intestinal mucosa heparin sodium salt from Sigma-Aldrich (9041-08-1), porcine intestinal mucosa DS from Celsus laboratories a gift from Malcolm Lyon - University of Manchester, porcine intestinal mucosa DS from Grampian Enzymes, bovine trachea CSA from Sigma-Aldrich (C9819) or shark cartilage CSC from Fluka (27043) was dissolved in 50 ml 50 mM Tris pH 7.2-8.0, 1-5 mM $CaCl_2$, 50 mM NaCl pH 7.5.

Depending on how quickly and how far the digest was to be completed, 0.025-0.1 international units of enzyme per 100 mg of GAG was added to the digest and incubated at 25-30 °C. The pH of the digest was maintained and more enzyme was added to maintain or speed up the digestion. CSA and CSC were digested with chondroitinase AC, DS was digested with chondroitinase B and heparin was digested with heparinase I, all enzymes were from Grampian Enzymes. Less enzyme was added to the CSC digests as chondroitinase AC seemed to have a higher turnover rate for that CSC compared to CSA. If digestion was poor, addition of 0.1-1 mg/ml BSA drastically improved enzyme activity, suggesting that some GAG batches were contaminated with proteases. The DS from Malcolm Lyon was found to be heavily contaminated by proteases and adding BSA was vital to achieve good digestion.

The digest was monitored by injecting 5-20 μl aliquots on a Superdex Peptide 10/300 column, with a flow rate of 0.5 ml/min in 100-200 mM ammonium bicarbonate or PBS, monitoring the 232 nm UV-trace. The incubation time varied from 4-72 hours. When sufficient digestion was achieved, the reaction was either heat inactivated at 90 °C for 5 minutes or immediately freeze-dried.

The freeze-dried material was dissolved in 2 ml ddH_2O, filtered through a 0.45 μm filter and injected on two serially connected 2.6x90 cm columns packed with BioGel P10 fine resin from BioRad (1504144). The flow rate was 0.4 ml/min in PBS or 150 mM NaCl and 5-8 ml fractions were collected as the 232 nm UV-trace was monitored on the BioCAD FPLC system. Ammonium bicarbonate was avoided as running buffer, although it has the advantage of being volatile, it miscolours the GAGs and makes them precipitate when concentrated.

To purify different specific species of heparin-derived tetrasaccharide and hexasaccharide, an analytical or semi-preparative AS-17 lonpac strong anion exchange column from Dionex was connected to the Waters HPLC system and a gradient from 100% buffer A to 100% buffer B was applied over 60 min at a 1 (analytical) or 4 (semi-preparative) ml/min flow rate, monitoring 232 nm UV-absorbance. Buffer A was dH_2O titrated with HCl to pH 3, buffer B was the same, but with 1 M NaC1. Samples were desalted by injecting up to 5 ml sample on a 1.6x30 cm GE Healthcare Sephadex G-25 superfine column with dH_2O as running buffer on the BioCAD FPLC system. If a small quantity of GAG needed desalting, the Superdex Peptide column was used instead. For preparation of large quantities of hexasaccharide for TEMPO-conjugation, the semi-preparative SAX column was connected to the BioCAD FPLC system and an automated method was setup to make 12 consecutive runs with the instrument autoinjector.

Chemically sulfated maltose and maltohexaose were kind gifts from Dr Craig Freeman, Division of Immunology and Genetics, John Curtin School of Medical Research, Australian National University, Canberra, Australia.

Fondaparinux was a kind gift from GlaxoSmithKline UK.

3.4 Biotinylation of GAG Oligosaccharides by Reductive Amination

Bovine lung heparin (Calbiochem 375093), dermatan sulfate (Celsus Laboratories) and chondroitin sulfate C (Fluka 27043) were digested with heparinase I, Chondroitinase B or Chondroitinase AC (Grampian Enzymes), respectively, and the decahexasaccharide was purified by gel filtration as described in section 3.3. Typically, $1x10^{-6}$ moles of decahexasaccharide was dissolved in 10% H_2O, 20% HOAc and 70% DMSO in a total volume of 200 μL. Thereafter, $5x10^{-6}$ moles of biotinamidohexanoic acid hydrazide (Sigma-Aldrich B3770) in DMSO and $1x10^{-4}$ moles of $NaCNBH_3$ in 50% DMSO / 50% H_2O were added to the reaction. The reaction mixture was incubated over night at 65 °C, preferably shaking, desalted on a 1.6x30 cm Sephadex G25 superfine column and lyophilised.

3.5 Affinity Chromatography

3.5.1 In-House Preparation of a Crude Heparin Affinity Column Linked via Free Amines

10 mg bovine lung heparin (Calbiochem 375093) dissolved in PBS and 2 mg EZ-Link NHS-PEG_4-Biotin (Pierce 21329) dissolved in PBS were mixed vigorously in an Eppendorf tube and incubated at room temperature for one hour. The sample was desalted on a 1.6x30 cm Sephadex G25 superfine column before loaded onto a GE healthcare HiTrap 1 ml streptavidin column.

3.5.2 In-House Preparation of DP16 GAG Affinity Columns

Biotin conjugated GAGs were prepared as described in section 3.4, dissolved in PBS and the 232nm UV absorbance was measured. Next, the GAGs were loaded onto a GE Healthcare 1 ml HiTrap Streptavidin column and the 232 nm UV absorbance of the flow through was measured again. The UV absorbance measurements were used to determine the conjugation and immobilisation efficiency, which varied between 71-88%.

The hyaluronan affinity column was manufactured by loading 5 mg of biotinylated hyaluronan from Sigma-Aldrich (B1557) onto at GE Healthcare HiTrap streptavidin column.

3.5.3 GAG Affinity Chromatography

Proteins were loaded onto the columns and eluted with a 50 CV gradient on the Akta system. Buffer A was 20 mM KPO_4 pH 7.5, 0.02% NaN_3. Buffer B was the same but contained 2 M NaCl.

For the hyaluronan affinity column, HGF-N was loaded in 20mM HEPES pH 7.5 and eluted with a 20 CV 0-1 M NaCl gradient on the BioCAD FPLC system.

3.5.4 Preparation of a HGF-N Affinity Column

Different conditions were explored, but in the final optimised reaction, 3 mg HGF-N in PBS was incubated with a tenfold excess of heparin DP10 and a twofold excess of NHS-PEG_4-Biotin from Thermo Scientific (21329) at room temperature for 30 minutes. To purify active HGF-N that was not biotinylated on the GAG binding site and has normal heparin affinity, the biotinylated protein was injected on a GE Healthcare HiTrap heparin column and eluted with a NaCl gradient. The high affinity biotinylated HGF-N fractions were buffer exchanged into PBS and loaded onto a GE Healthcare HiTrap streptavidin column. A_{280} was measured before and after loading the biotinylated protein onto the column and it was determined that $1.8x10^{-7}$ mole HGF-N was bound to the finalised HGF-N affinity column.

3.5.5 HGF-N affinity chromatography

To perform HGF-N affinity chromatography on a mixture of heparin-derived DP6 species, $2\text{-}4x10^{-7}$ mole of hexasaccharide was loaded onto the column in PBS and eluted with a NaCl gradient from 100 mM NaCl to 1 M NaCl buffered with KPO_4 pH 7.2 in 80 minutes with a flow rate of 1 ml/min. The 232 nm UV trace was monitored and fractions collected on the Waters HPLC system.

3.5.6 High Resolution Heparin Affinity Chromatography of truncated HGF-N

Protein samples were loaded on a 1 ml heparin HiTrap column from GE healthcare connected to the Akta FPLC system. The gradient was as follows: 0-35 % B in 4 CV, 35-75 % B in 30 CV, wash with 100 % B and equilibrate.

3.6 Characterisation of Dimeric HGF-N

To investigate what caused the protein to elute as two peaks and what was different between the two forms, high affinity protein was purified by standard heparin chromatography and diluted 1/5 with buffer A. Half of the high affinity protein was incubated at 37° C for 2.5 hours and the remaining half was incubated at 4° C for 2.5 hours. After the incubation, both samples were analysed by heparin affinity chromatography again.

To characterise the oligomeric state of the high and low affinity forms of HGF-N, gel filtration was carried out at 4° C on the Akta system, using a XK16/60 Superdex HiLoad column in PBS, collecting 2 ml fractions. HGF-NK-G was used as a 20.5 kDa standard, representing the size of dimeric HGF-N. The position of the conductivity peaks of the high salt buffer, in which both protein samples were dissolved, confirmed that samples were injected in the same manner. It was found that the high affinity form was likely a dimer. The dimeric protein was isolated from the gel filtration column, incubated at 37° C for 2.5 hours, spin concentrated and injected onto the gel filtration column again. This time the protein eluted as a monomeric species.

For NMR characterisation, heparin chromatography was carried out on ^{15}N labelled protein to isolate the high affinity dimeric form of HGF-N. The high affinity fractions were kept on ice or at 4 °C as they were spin concentrated and buffer exchanged into NMR buffer. The final NMR sample was put into the AVA800 spectrometer and the dissociation of the dimeric high affinity form of HGF-N into its

normal affinity monomeric form was studied by collecting 6 consecutive 5-15 minute SOFAST ^{1}H-^{15}N HSQCs followed by 57 minute long ^{1}H-^{15}N HSQC experiments.

3.7 Gel filtration Studies of HGF-N and HGF-NK-G Complexed With Different GAGs

Gel filtration was performed to investigate the oligomeric state of HGF-N and HGF-NK-G in the presence of GAGs. Before the proteins were analysed by gel filtration, they were dialysed into PBS buffer and mixed with an equimolar amount of GAG dissolved in dH_2O. The 400-800 μL protein + GAG sample was incubated for 5 minutes at room temperature before being analysed by gel filtration. A XK16/60 Superdex HiLoad column from GE Healthcare was equilibrated with PBS at room temperature and used for all the experiments. The column was attached to the Akta FPLC system with a 1 ml/min PBS flow rate, monitoring the A_{280} trace and collecting 2 ml fractions. After all experiments were finished, the collected protein fractions were repurified by heparin chromatography to remove the GAGs and the protein was reused for other experiments.

3.8 Analytical Ultracentrifugation

Analytical Ultracentrifugation experiments were carried out with the help of Alexandra Solovyova, University of Newcastle.

The sedimentation velocity experiments were performed on a Beckman Coulter ProteomeLab XL-I analytical ultracentrifuge using both UV absorbance at 280 nm and interference optics. All AUC runs were carried out at 48,000 rpm and at 30 °C. The sample volume was 400 μl and the sample concentrations ranged between 0.25 and 0.95 mg/ml. The protein GAG ratios were 1:1. The heparin DP16 used in the AUC runs was purified as described, but reapplied to the BioGel P10 column once again and only the middle of the peak was used to produce a more homogeneous GAG sample.

The partial specific volumes ($\bar{v}$) for the proteins were calculated from the protein amino acid sequence, using the program SEDNTERP [224] and extrapolated to the experimental temperature following the method of Durchschlag [225] in (3.1).

$$\bar{v}_T = \bar{v}_{298.15} + 4.25 \times 10^{-4}(T - 298.15) \qquad \textbf{(3.1)}$$

The partial specific volume value of Fondaparinux and heparin DP16 was assumed to be 0.5 ml/g [226], [227]. The partial specific volumes of the protein-GAG complexes were calculated using (3.2).

$$\bar{v}_{complex} = \frac{M_{protein}\bar{v}_{protein} + M_{GAG}\bar{v}_{GAG}}{M_{protein} + M_{GAG}} \qquad \textbf{(3.2)}$$

The distribution of sedimenting material was modelled as a distribution of Lamm equation solutions [228] where the measured boundary a(r,t) was modelled as an integral over differential concentration distributions c(s) using (3.3).

$$a(r,t) = \int c(s)\chi(s,D,r,t)ds + \varphi \qquad \textbf{(3.3)}$$

where φ is the noise component, r is the distance from the centre of rotation diffusion, D is the diffusion coefficient and t is time. The expression χ(s, D, r, t) denotes the solution of the Lamm equation [229] for a single species by finite element methods [230].

The integral for (3.3) was implemented in the program SEDFIT 11 Beta (http://www.analyticalultracentrifugation.com) and solved numerically by discretisation into a grid of 200 sedimentation coefficients for both absorbance and interference data. The best-fit concentrations for each plausible species were calculated as linear least square fits. The sedimentation velocity profiles were fitted using a maximum entropy regularisation parameter of p = 0.95. The weight average

sedimentation coefficient was calculated by integrating the differential sedimentation coefficient distribution using (3.4) [231].

$$s_w = \frac{\int c(s)ds}{\int cds} \qquad \textbf{(3.4)}$$

Sedimentation coefficients were extrapolated to zero concentration and converted to 20°C H_2O standard conditions. The diffusion coefficient (D) corresponding to each sedimentation coefficient value was estimated from a weight-average frictional ratio $(f/f_0)_w$ [231] as in (3.5).

$$D(s) = \frac{\sqrt{2}}{18\pi} kTs^{-1/2}[\eta(f/f_0)_w]^{-3/2}[(1-\bar{v}\rho)/\bar{v})]^{-1/2} \qquad \textbf{(3.5)}$$

The integration of the mass distribution c(M) with a similar approach to (3.4) gave the weight average molecular mass of the solute.

A two-dimensional size-and-shape distribution model ($c(s,f/f_0)$) was applied in order to precisely determine the value of molecular mass from sedimentation velocity data [232]. In this model a differential distribution of sedimentation coefficients and frictional ratios (f/f_0) was defined using (3.6).

$$a(r,t) = \int\int c(s,f_r)\chi(s,D(s,f/f_0),r,t)dsdf_r \qquad \textbf{(3.6)}$$

where all symbols are the same as in (3.3).

In this model, the sedimentation coefficient and friction ratio, i.e. mass, for sedimenting species were fitted as independent variables. The numerical solutions of

(2.6) were obtained in a discrete grid of 100 sedimentation coefficients and 20 values of f/f_0.

The HYDROPRO 7 [233] software was used to calculate the protein sedimentation coefficient based on their atomic coordinates using the 2HGF and 1GMN PDB files.

3.9 Isothermal Calorimetry

ITC data was collected on a Microcal Auto VP-ITC at 30 °C, the autosampler was maintained at 4 °C. The cell volume was 1.4 ml and the syringe volume was 280 μL. The ligand was titrated by 44 x 6.4 or 88 x 3.2 μL injections lasting 12.8 seconds with a 180 second delay between injections. Protein and GAG concentrations were carefully determined by A_{280} and A_{232}, respectively, in a glass cuvette on SpectraMax M5 multimode plate reader from Molecular Devices. The protein concentrations varied between 75 and 95 μM in PBS and the GAG concentrations were 630-1100 μM in PBS. The following control titration experiments were performed and subtracted: buffer into protein, ligand into buffer and buffer into buffer. The data was analysed in the Microcal version of Originlab 7E.

3.10 NMR Monitored Thermal Denaturation

The AVA600 spectrometer and the multi_zgvt command were used to collect a series of ^{15}N-^{1}H HSQC spectra and 1D ^{1}H spectra using Double-Pulse-Field-Gradient-Spin-Echo for water suppression. Spectra were collected at 30 °C, 50 °C, 60 °C, 70 °C, 75 °C, 80 °C, 85 °C and 90 °C with a temperature equilibrium period of 800 s between each experiment. A 105 μM ^{15}N labelled sample of HGF-N was used to collect spectra in the presence of ten times excess heparin DP10 and a 56 μM sample was used to collect spectra for the free protein.

3.11 NMR Titrations

Spectra were acquired on the AVA800 or BIO600 Spectrometer at 30 °C. The samples were dialysed using a 1000 MWCO tube-o-dialyzer (WebScientific 786-143) at 4°C for approximately 12 hours against 100 mM NaCl, 20 mM potassium phosphate, pH 6.8. The starting sample volume was 600 µl and protein concentrations were 50 - 100 µM containing 5-10 % D_2O and 0.001-0.002 % w/v sodium azide. The pH was monitored during the titration, but normally no addition of acid or base was needed.

GAG oligosaccharides used in titrations were desalted, quantified by A_{232}, freeze-dried and re-dissolved in H_2O, producing very concentrated stock solutions that were then diluted so that no more than 1 µL was added to the sample at each titration point.

To estimate the K_Ds of the interactions, CcpNmr Analysis was used to fit titration data according to (3.7).

$$y = A\left((B+x) - \sqrt{(B+x)^2 - 4x}\right) \qquad \textbf{(3.7)}$$

Where $y = ChemShif\,tChange$,

$A = \frac{MaxChemShif\,tChange}{2}$ and

$B = 1 + \left(\frac{K_D}{[protein]}\right)$

The K_D was then determined in Open Office Calc 2.3 by solving the equation $K_D = (B-1) \times [protein]$ for residues with large CCSDs fitting well to the equation and the average of these residues was then used as the K_D.

The combined chemical shift perturbation was expressed as in (3.8).

$$\sqrt{(\Delta\delta H)^2 + \left(\frac{\Delta\delta N}{5}\right)^2} \qquad \textbf{(3.8)}$$

where ΔδH and ΔδN are the chemical shift changes in the proton and nitrogen dimension, respectively, between the unbound and bound state in ppm.

3.12 Reductive Amination to Conjugate 4-Amino TEMPO to Fully Sulfated Heparin Hexasaccharide

Fully sulfated heparin hexasaccharide was purified by standard SAX chromatography and desalted on a Sephadex G25 superfine column. Typically, five hundred nanomole mole hexasaccharide was lyophilised and dissolved in 10% dH_2O, 70% DMSO and 20% acetic acid, 1 M 4-amino TEMPO in DMSO was added to thirtyfold molar excess and 5 M $NaCNBH_3$ in 50% DMSO/dH_2O was added to three hundredfold excess. The reaction was incubated at 65 °C over night, preferably with shaking. As a last step, the reaction was desalted on a Sephadex G25 superfine resin and purified by SAX chromatography.

3.13 Titration of TEMPO Conjugated Heparin Hexasaccharide into ^{15}N-Labelled HGF-N Monitored by NMR

A 75 μM ^{15}N labelled sample of HGF-N was prepared in 550 μl of the standard NMR buffer and all experiments were carried out on the AVA800 spectrometer. NMR data was collected at the following TEMPO-heparin DP6:protein ratios: 0:1, 0.5:1, 1:1 and 1.8:1. A 1D ^{1}H spectrum was collected at each titration point, using Double-Pulse-Field-Gradient-Spin-Echo for water suppression [234]. A ^{1}H-^{15}N HSQC spectrum was collected to determine the chemical shift perturbation for each titration point. To determine the T_1 relaxation times of NH protons at each titration point, a pulse sequence that combines inversion recovery with a ^{1}H-^{15}N HSQC was used [235]. T_1 values were measured at each titration point using the following

relaxation delay times: 40, 100, 250, 600, 800 and 1200 ms. The t_1 and t_2 acquisition times were 92 and 316 ms respectively. The spectral width was 30 ppm in the nitrogen dimension and 14 ppm in the proton dimension. For each t_1 increment, 16 scans were collected and the total experiment time needed to acquire one relaxation point at a given protein:GAG ratio was approximately 12 hours.

To determine T_1 values, spectra were processed in Azara and imported into CcpNmr Analysis 2.1.2. Peaks were assigned, peaklists copied to all delay times and the experiment series for each titration point were fitted to the equation $A\left(\frac{1}{2} - e^{-Bx}\right)$ with jiggling error method. The contribution of the spin label to the relaxation of NH protons was determined by subtracting the relaxation rate $R_1 = \frac{1}{T_1}$ of the uncomplexed HGF-N from the relaxation rate measured using the 1:1 TEMPO-heparin DP6: HGF-N complex. HSQC shift perturbations were quantified using Analysis and equation (3.8). The relaxation and chemical shift perturbation data were plotted in Open Office Calc 3.1.1. Finally, crosspeaks with a shift dist ≥0.09 ppm and with a delta R_1 ≥ 0.00035 ms for the 1:1 TEMPO-heparin DP6:HGF-N titration point were mapped onto protomer C of the 1GMO pdb file using UCSF Chimera 1.4.

4 Results and Discussion

4.1 Design and Comparison of Different HGF Protein Constructs

Several different protein constructs are used in this thesis and references are made to protein constructs used by other research groups. These protein constructs are summarised in Table 4.1.

As explained in section 1.2.1 and Figure 1.1, HGF consists of the following six domains from the N- to the C-terminus: N, K1, K2, K3, K4 and finally the SPH domain. In this project, HGF-N denotes the N domain of HGF, HGF-NK, the N and the K1 domains, HGF-NK-C denotes the NK fragment that was cloned in this project and HGF-NK-G, the NK fragment that was originally cloned in Cambridge by the Gherardi research group.

HGF-N is the same protein construct that was expressed in *E. coli* and used for determining the NMR structure of the HGF N domain by Zhou *et al* [21]. This protein construct was chosen to allow for easy transfer of the published assignment of ^{1}H-^{15}N crosspeaks. The N-domain contains the GAG binding site of HGF and the main purpose of this protein construct was to use it for NMR titrations of GAG oligosaccharides.

HGF-K was expressed with an intention to investigate if the K domain contributes to GAG binding, but this protein construct was never used in any experiments.

HGF-NK-C was expressed to allow for comparison with the HGF-N and HGF-NK-G protein constructs. The latter protein construct has been extensively studied by the Gherardi research group in Cambridge, for example this protein construct was used to determine the HGF-NK-G – heparin co-crystal structure [14]. Initially, HGF-NK-C was the only source of NK protein before HGF-NK-G was obtained from a collaborator.

Table 4.1 Table summarising different HGF constructs, their N-termini, PDB files and protein constructs. The red Q indicates the first residue of the native protein. Bold letters represent residues which coordinates can be seen in the PDB files. Potentially phosphorylated residues are indicated with ^{P}.

Name in this thesis	N-terminus	Residues	PDB file	Reference
Native HGF (ExPASy P14210)	QRKRRNTIHEFKKS…	32-728	N/D	Section 1.2
HGF-NK-G	(EA)(EA)YVEGQRKRR**NT IHEFKKS…**	30-210	1GMO	Section 4.5
HGF-N	GQRKRRNTIHEFKKS…	31-127	2HGF	Section 4.2.1
HGF-NK-C	GQRKRRNTIHEFKKS…	31-208	N/D	Section 4.2.3
1BHT	SDYKDDDDKLRRNTIHE QRK**RRNTIHEFKKS…**	35-210	1BHT	Section 4.3
HGF-N-PO$_4$	GQRKRRNT^{P}IHEFKKS^{P}**…**	31-127	N/D	Section 4.2.1
HGF-N-Δ6	NTIHEFKKS…	37-127	N/D	Section 4.2.1
HGF-N-Δ5	RNTIHEFKKS…	36-127	N/D	Section 4.2.1

4.2 Expression, Purification and Characterisation of Protein Constructs

4.2.1 HGF-N

The cloned version of HGF-N is 97 amino acids, has a molecular weight of 11376.2 Da and a theoretical pI of 10.03. The first residue of the construct is 31G. The very first residues of the native protein (residues 1-32) belong to the signal peptide that is cleaved after translation and the first residue of the native HGF protein is 32Q.

Pichia transformants expressing HGF-N were identified by miniscale protein expression in Falcon tubes and analysed by SDS-PAGE (Figure 4.1). The first three clones from the left (A-C) grew on 300 μg/ml YPDSZ plates, the remaining three (D-F) grew on 100 μg/ml YPDSZ plates. HGF-N can be clearly seen as a double band at about 15 kDa. It is also clear that the clones with higher Zeocin resistance express more protein compared to the low Zeocin resistance clones, as expected.

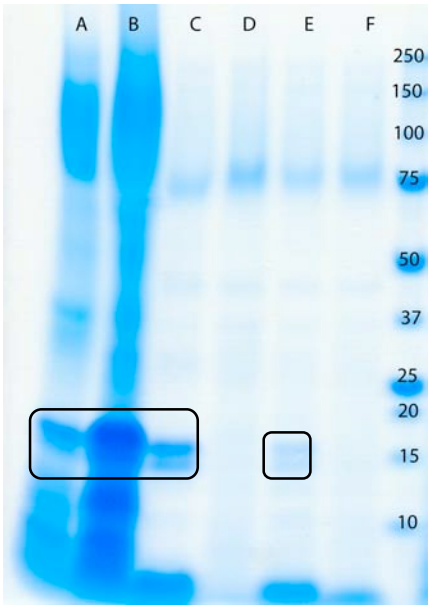


Figure 4.1 SDS-PAGE analysis of a mini-scale expression of HGF-N. The supernatant of a 5 ml Falcon-tube cell-culture was loaded onto the gel. Lanes A-F are different clones. The lane furthest to the right is the protein ladder and the numbers indicate the kDa weight of the proteins in the ladder. The rectangles indicate HGF-N expression. The high background in lane B is likely due to that the pipette tip touched the supernatant when preparing the sample and intracellular proteins from lysed cells contaminated the SDS-PAGE gel lane. Clone C was chosen for large scale protein production.

An unlabelled version of HGF-N was produced for general experiments as well as a ^{15}N-labelled version for NMR experiments. ^{15}N labelled protein was expressed in a similar manner but on 1-litre scale. The protein was initially purified on a heparin-Sepharose resin and eluted as a peak with a shoulder at 15-22 minutes (Figure 4.2). Unspecific binding of *Pichia* proteins was also visible in the chromatogram. A 5-litre unlabelled fermentation of HGF-N yielded approximately 160 mg of heparin affinity purified protein.

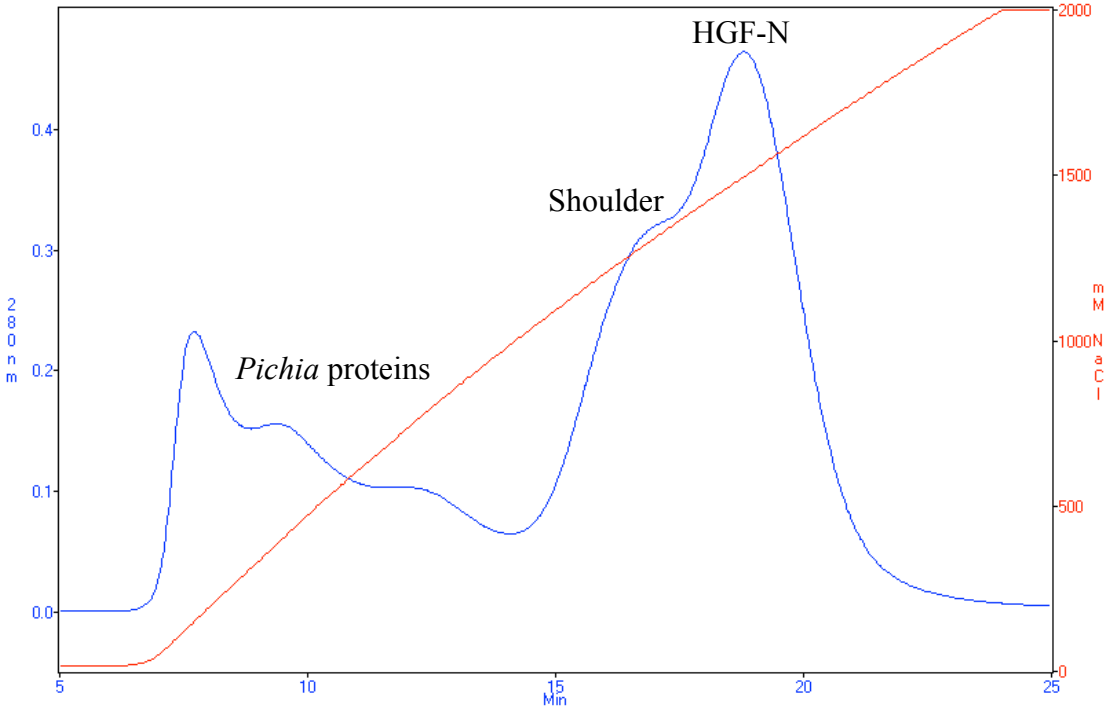


Figure 4.2 Elution of ^{15}N labelled HGF-N from heparin-Sepharose fast flow beads. The blue line is the A_{280} trace, the red line is the conductivity trace.

To assess the purity of the heparin affinity-purified protein, fractions were analysed by SDS-PAGE (Figure 4.3).

As evident from Figure 4.2 and Figure 4.3 the expression of HGF-N did not produce a homogeneous protein and further purification was necessary. SDS-PAGE analysis showed two bands of similar size. The lower molecular weight band eluted first and the higher molecular weight band eluted later from the heparin-Sepharose resin (data not shown), implying that the higher molecular weight bands had higher heparin affinity. Similar molecular weights of the bands suggested that gel filtration would not be successful, therefore MonoS strong cation exchange chromatography was carried out (Figure 4.4).

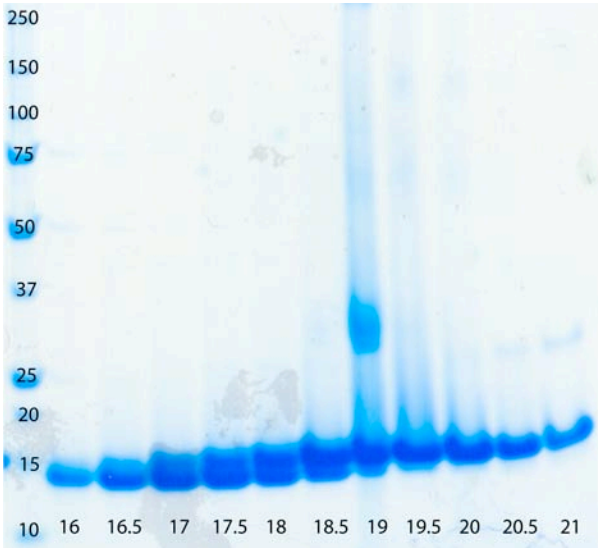


Figure 4.3 SDS-PAGE analysis of fractions eluting between 16 and 21 minutes from the heparin-Sepharose affinity chromatogram in Figure 4.2. The lane furthest to the left is the protein ladder and the numbers indicate the kDa weight of the proteins in the ladder. The numbers at the bottom of the gel indicates elution time in Figure 4.2. The band at 30 kDa in lane 19 is likely a contaminant from the SDS-PAGE analysis.

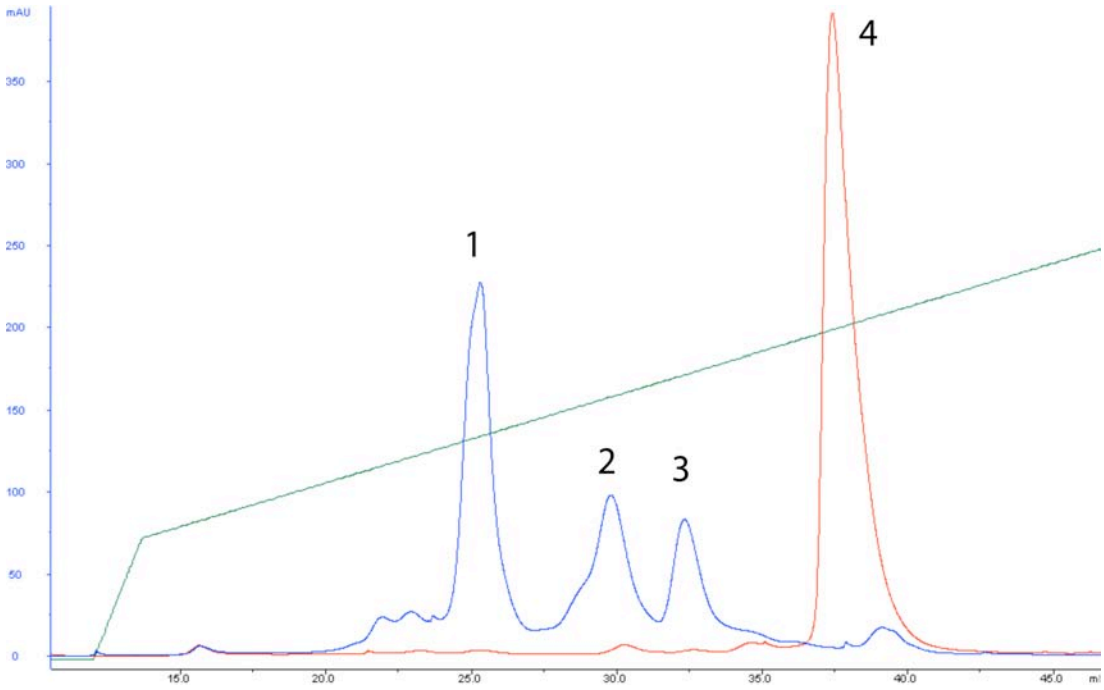


Figure 4.4 MonoS purification of HGF-N. The blue trace is a fraction from the first shoulder (17 minutes) from the heparin-Sepharose column in Figure 4.2, the red trace is from a late fraction of the major peak (21 minutes).

MALDI-TOF mass spectrometry confirmed that the red peak 4 in Figure 4.4 is a pure species with the correct MW of native HGF-N (Figure 4.5). Preliminary MALDI-TOF results (data not shown) confirmed that peak 1 and 2 in Figure 4.4 contained truncated proteins and that the third peak had a posttranslational modification, possibly a phosphorylation. As the different MonoS peaks exhibited different heparin affinities, it was relevant to investigate in detail what the differences between the peaks were.

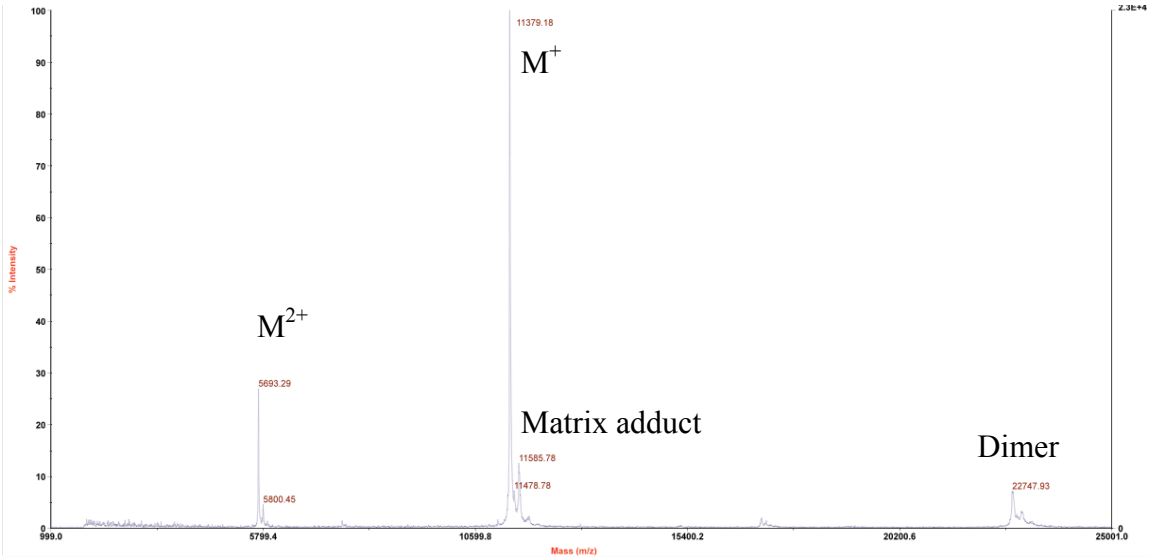


Figure 4.5 MALDI-TOF spectrum of MonoS purified HGF-N. No internal standards were used to calibrate the spectrum. The experimental mass was determined to 11378.2 Da, the theoretical mass of 11372.2 Da. As no internal standards were used in the MALDI spectrum, the experimental mass is within acceptance criteria and a more accurate mass was determined by FTICR-MS.

To confirm and further clarify the MALDI-TOF results, one and two-dimensional FTICR mass spectrometry was employed (Figure 4.6 and Figure 4.7). Data was collected with the help of Jenna Scotcher, SIRCAMS.

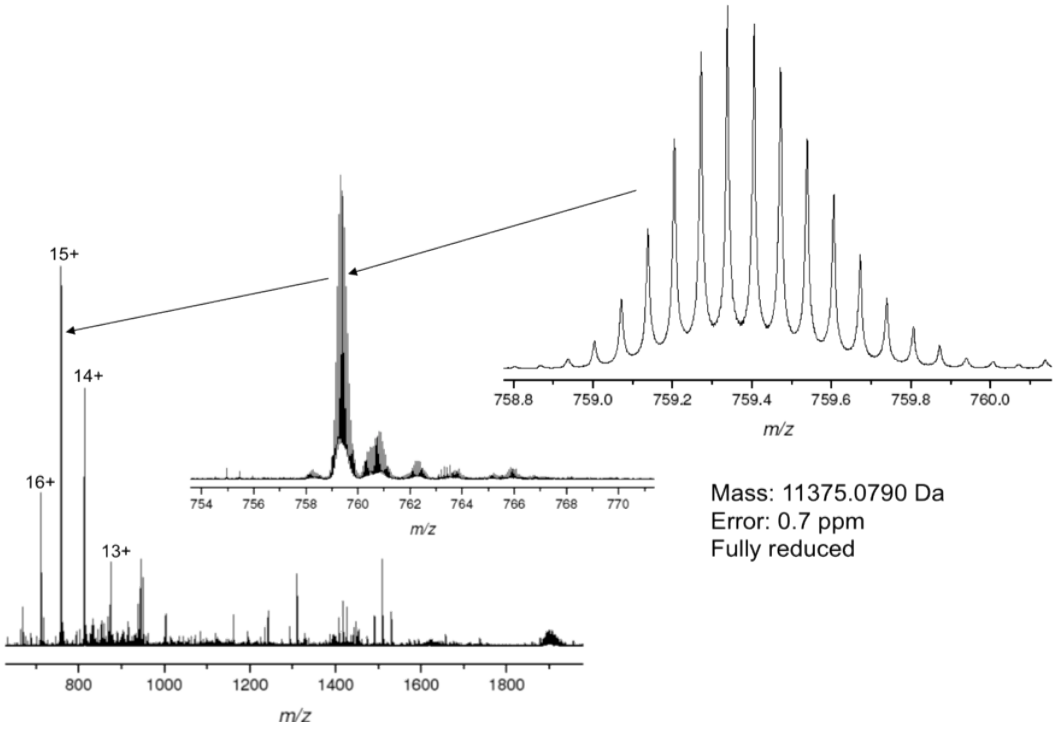


Figure 4.6 FTICRMS spectra of HGF-N. The 15^{+}-charge state was used for accurate mass determination.

Accurate mass determination and isotope simulation of the non-native proteins purified by MonoS chromatography clearly indicated that the proteins had N-terminal truncations and that one version was phosphorylated. To further confirm the results, electron capture dissociation spectra were collected (Figure 4.7). Sequence coverage was achieved by both the c and z ions.

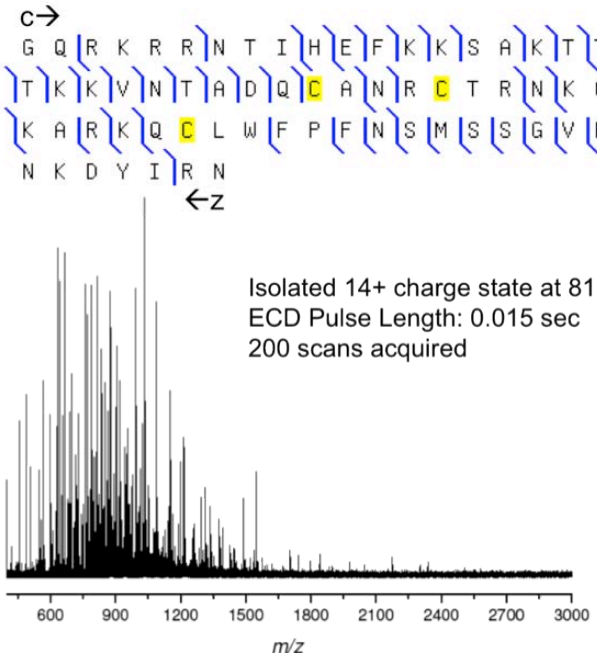


Figure 4.7 Electron capture dissociation FTICR mass spectrum of native HGF-N.

The FTICR mass spectrometry analysis results are summarised in Table 4.2.

Table 4.2 Summary of FTICR MS mass determination and isotope simulation data. See Figure 4.4 for the chromatogram peak numbering.

Peak	**Mass** (Da)	**Error** (ppm)	**Lacking residues**	**NMR HSCQC**	**Name**
Peak 1 Blue	10593.5	0.3	- GQRKRR	Lacks 32Q, binding site shift differences	HGF-N-Δ6
Peak 2 Blue	10749.6	1.6	- GQRKR	Lacks 32Q, binding site shift differences	HGF-N-Δ5
Peak 3 Blue	11456.0	1.5	+ PO_4	Binding site shift differences	HGF-N-PO_4
Peak 4 Red	11375.1	0.7	-	Reference	HGF-N

^{1}H-^{15}N HSQC NMR spectra were collected for both the native and truncated HGF-N proteins. In general, the crosspeaks of the truncated HGF-N-Δ6 and HGF-N-Δ5 proteins overlaid well with HGF-N, but amino acids close to the N-terminus and those of the heparin binding site α2-helix exhibited chemical shift changes. Crosspeaks for residues 32Q and 38T were missing, crosspeaks for residues 39-41 were shifted and residues 70-75 were also shifted, as they are close in space to the N-terminus (Figure 4.8).

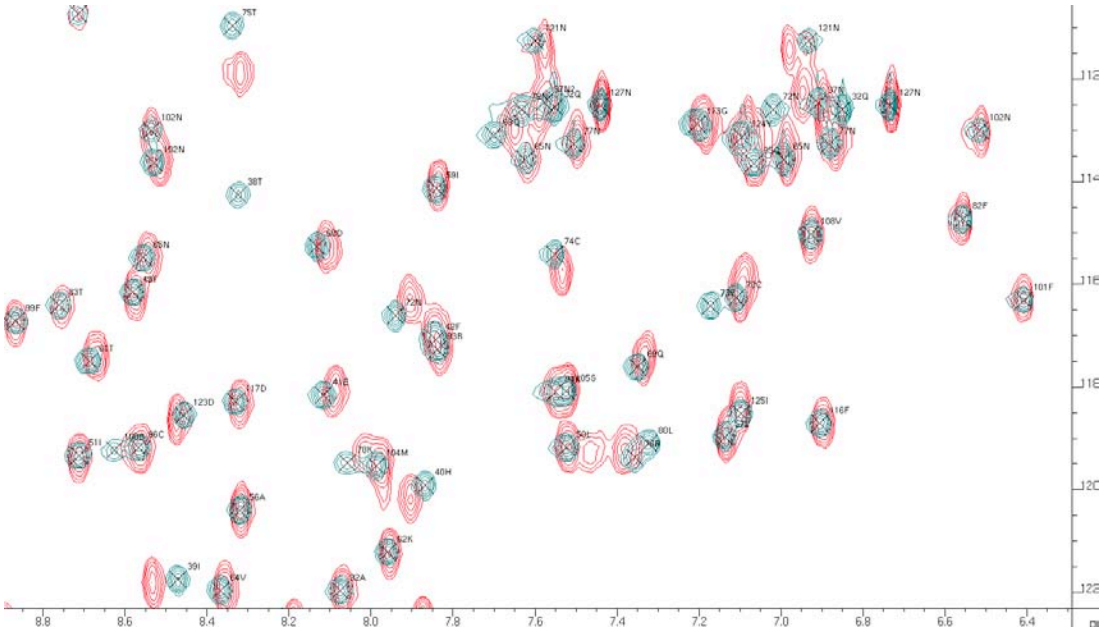
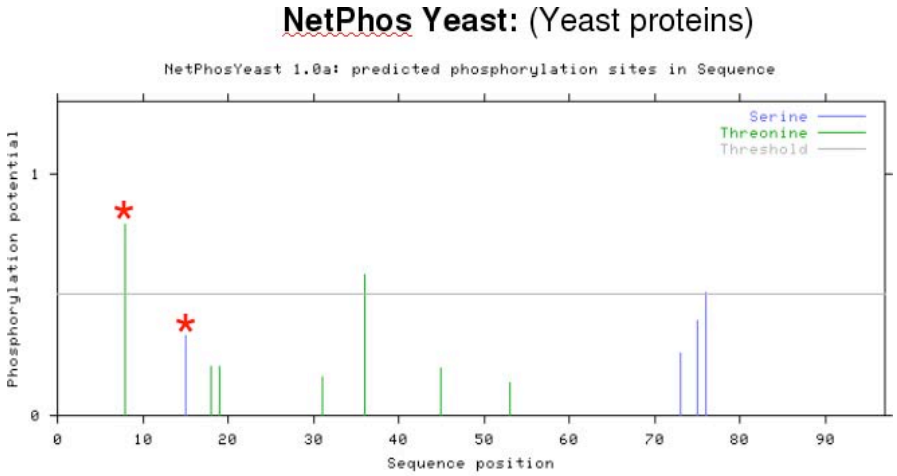


Figure 4.8 Partial ^{15}N-^{1}H HSQC spectrum of HGF-N-Δ6 in red overlaid with the ^{15}N-^{1}H HSQC spectrum of HGF-N in green. The crosspeaks in the blue spectrum are somewhat elongated due to fewer increments in the nitrogen dimension.

Phosphorylation by *P. pastoris* has not been observed in our laboratory before and is not widely mentioned in the literature. FTICR ECD mass spectrometric analysis showed that either T38 or T38 and S45 could be phosphorylated. Phosphorylation site prediction with NetPhos 2.0 and NetPhos Yeast (http://www.cbs.dtu.dk/services/NetPhos/) predicted both T38 and S45 as possible phosphorylation sites (Figure 4.9). To date no phosphorylation sites of HGF have been mapped (http://www.phosphosite.org, http://phospho.elm.eu.org, 2010-05-25).

The phosphorylations are interesting as they are a possible way to regulate the HGF-GAG interaction and thus the whole signalling cascade as discussed in section 1.2.



NetPhos 2.0: (Eukaryotic proteins)

Sequence	8	KRRNTIHEF	0.994	*T*
Sequence	15	EFKKSAKTT	0.900	*S*

Figure 4.9 NetPhos Yeast (upper histogram) and NetPhos 2.0 (lower table) phosphorylation site prediction results. The two mapped phosphorylations (T38 and S45) are marked with red asterisks for NetPhos Yeast (upper histogram), for NetPhos 2.0 (lower table) the score is given. The higher the score, the higher the likelihood of the specific residue being phosphorylated, 1 is the highest score.

The 5 or 6 amino acid truncation of some of the HGF-N protein is due to faulty Kex2 cleavage during secretion of the protein in *P. pastoris*. When the proteins were cloned into the PPICZAlpha vector, the α-factor signal sequence was used to direct the expressed protein for secretion. The Kex2 enzyme in *P. pastoris* cleaves the α-factor signal sequence from the expressed protein as it is being secreted. The Kex2 cleavage takes place between Arg and X in the sequence X - X - Lys/Arg - Arg ^ - X sequences, where ^ indicates the cleavage site. Kex2 has a complex and promiscuous preference for the X residues [236] and usually, the X amino acids following after the cleavage site are glutamic acid and alanine for proteins cloned into a

PPICZAlpha vector, but in this thesis work, the proteins were cloned without these residues.

In conclusion, Kex2 cleaves at KR or RR sequences in the secretion signal sequence and HGF-N has this sequence in the N-terminal part of the protein. This confuses Kex2 and two cleavage sites are created. Kex2 does not recognize the RK sequence present in the N-terminus, only KR and RR as can be seen in (Figure 4.10).

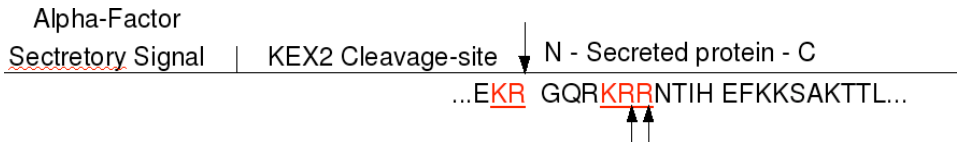


Figure 4.10 The first native kex2 cleavage site in the vector α-mating factor and the two new cleavage sites produced by the KRR sequence in the protein sequence.

4.2.2 HGF-K

The cloned version of HGF-K contains 83 amino acids, has a molecular weight of 9427.6 Da and a theoretical pI of 8.29.

The protein was first expressed in Falcon tubes (Figure 4.11) and then fermented in a bench-top fermenter in the same manner as HGF-N.

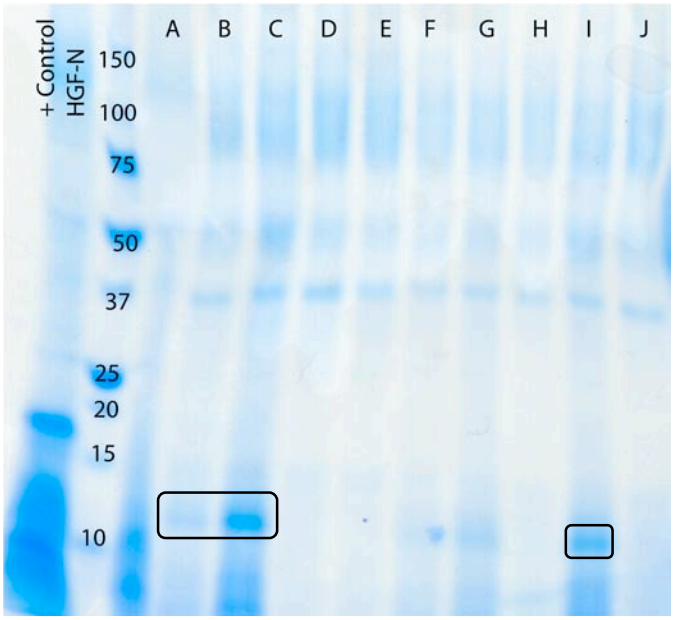


Figure 4.11 SDS-PAGE gel of mini-scale supernatants of HGF-K. The lane furthest to the left is a positive control of a clone expressing HGF-N. The second lane from the left is a protein ladder and numbers indicate the kDa weight of the proteins in the ladder. Lanes A-J are different clones. A clear band at 13 kDa in lanes B, C and I suggests protein expression of HGF-K as indicated by the rectangles. Clone C was chosen for large scale protein production.

Due to the lower pI and heparin binding of HGF-K compared to HGF-N and HGF-NK-C, a SP-Sepharose resin was used instead of heparin-Sepharose and the supernatant was titrated to pH 4.2 before loading to achieve good binding to the resin. The lower pH caused higher binding of secreted *Pichia* proteins to the resin as evident from a large A_{280} peak before the main HGF-K peak during the elution of HGF-K from the SP- Sepharose resin (Figure 4.12).

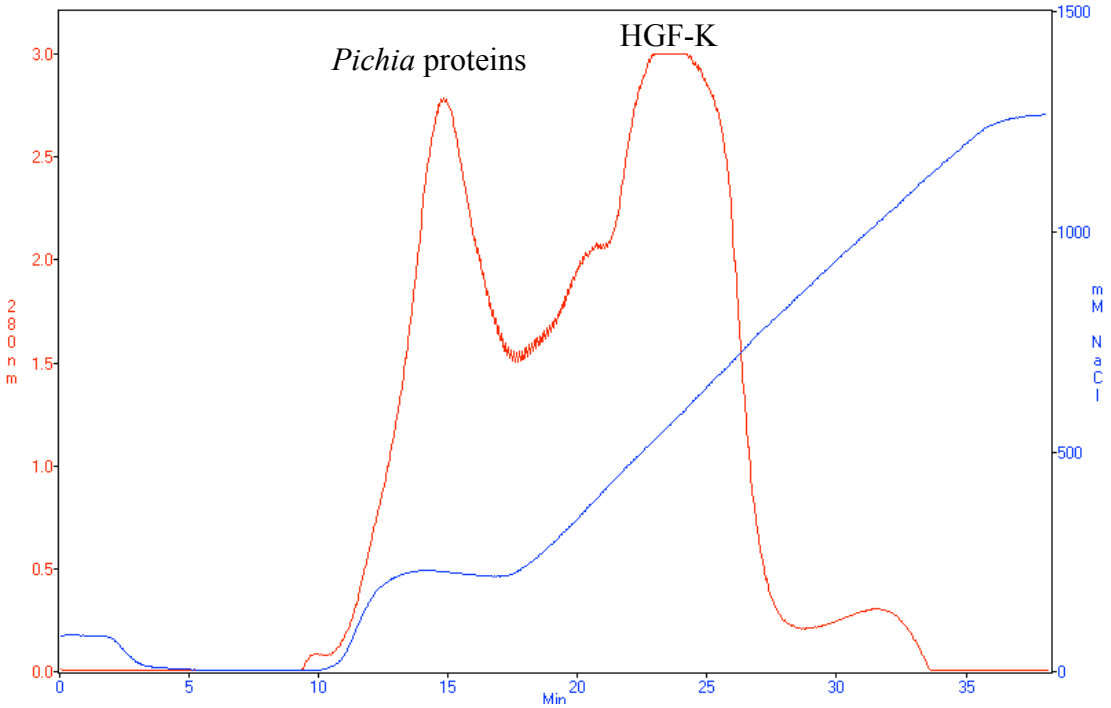


Figure 4.12 Chromatogram of the elution of ^{15}N labelled HGF-K from a SP-Sepharose resin. The UV-trace is shown in red and the conductivity is shown as mM NaCl in blue. The first peak corresponds to *Pichia* proteins and the second peak to HGF-K.

Higher A_{280} values observed when eluting HGF-K compared to when eluting HGF-N from the SP-Sepharose resin (Figure 4.12 vs Figure 4.2) indicates higher expression of HGF-K, despite the fact that the HGF-K extinction coefficient is twice that of HGF-N. The first eluting peak contains secreted *Pichia* proteins, as evident from the gel in Figure 4.13. A longer wash step would have given better separation between the *Pichia* protein and HGF-K peaks.

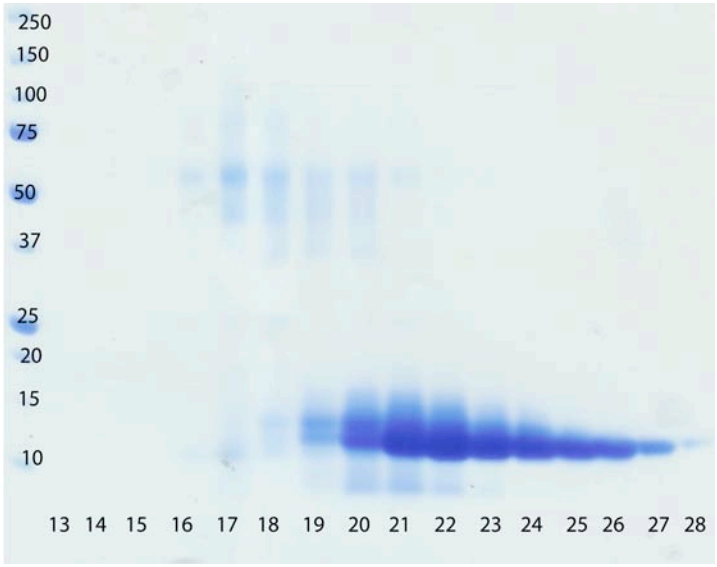


Figure 4.13 SDS-PAGE gel of the major HGF-K peak fractions in Figure 4.12. Each lane represents one fraction. The first lane from the left is a protein ladder and numbers indicate the kDa mass of the proteins in the ladder.

The beginning of the HGF-K peak at 22-27 min (Figure 4.12) contains a double band (Figure 4.13), which implied that either the later eluting protein had been truncated or that the earlier eluting protein had a post-translational modification. *P. pastoris* often glycosylates secreted proteins with mannose sugar units that can be highly branched. MALDI-TOF analysis suggested that the protein was not homogeneous (Figure 4.14).

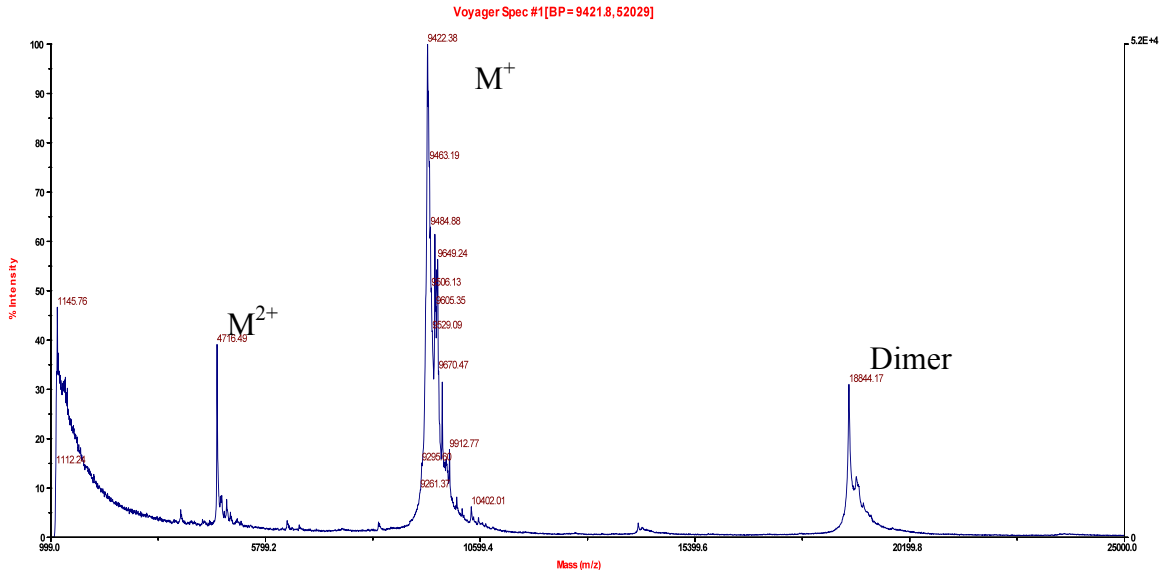


Figure 4.14 MALDI-TOF spectrum of fraction 22 of the SP-Sepharose purified HGF-K in Figure 4.13.

To purify the protein further, gel filtration was investigated, using fractions 24-28 from Figure 4.12, the chromatogram can be seen in Figure 4.15.

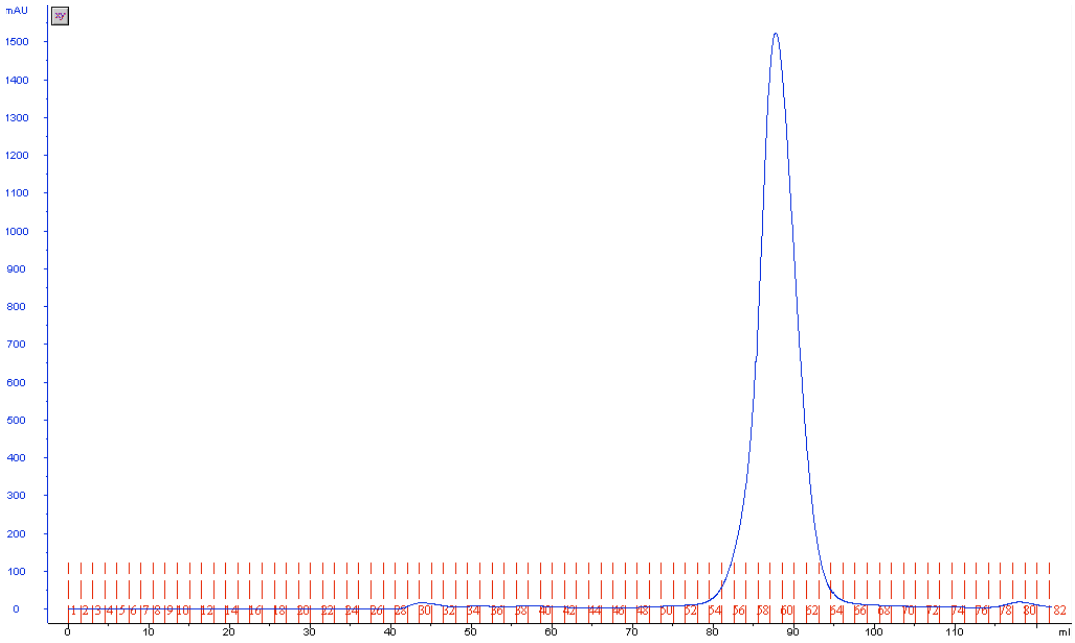


Figure 4.15 Gel filtration chromatogram of ^{15}N labelled HGF-K.

As can be seen in Figure 4.15, the gel filtration chromatogram only contains one main symmetrical peak. SDS-PAGE analysis (Figure 4.16) of the fractions showed the presence of higher molecular weight species in the early part of the peak.

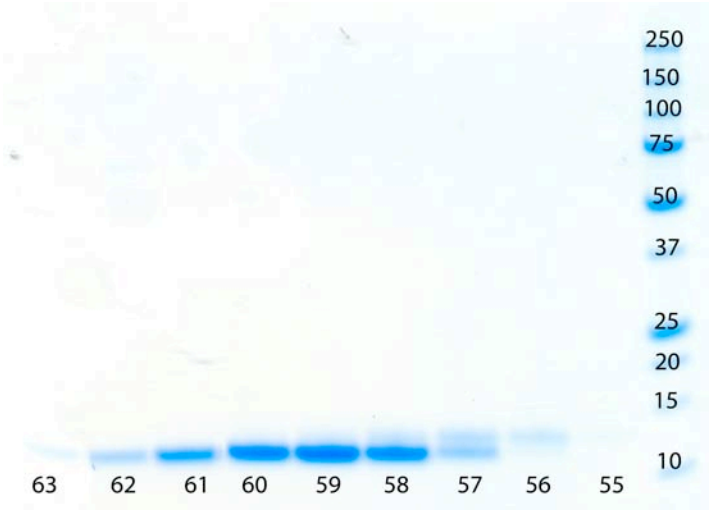


Figure 4.16 SDS-PAGE analysis of peak fractions in Figure 4.15. The protein ladder with its molecular weights is in the lane furthest to the right. The other lanes are chromatography fractions from Figure 4.15. A double band can be observed in lanes 56-58 and a single band in lanes 59-63.

Fraction 59 to 63 in Figure 4.16 were pooled and concentrated to collect a ^{15}N-^{1}H HSQC spectrum (Figure 4.17).

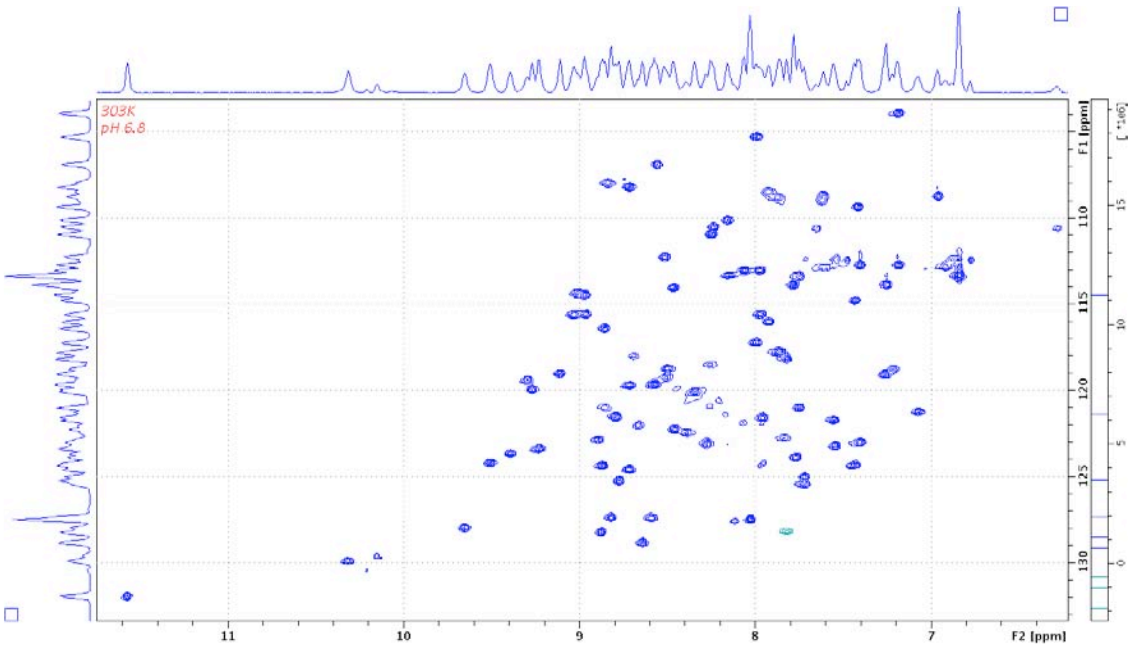


Figure 4.17 ^{15}N-^{1}H HSQC spectrum of ^{15}N labelled HGF-K.

The HGF-K ^{15}N-^{1}H HSQC spectrum is characteristic of a folded protein with well dispersed, strong peaks. The low intensity peaks in the middle of the spectrum are likely due to the protein not being purified to homogeneity.

No further work was done on HGF-K. If more experiments were to be performed, the possible modification would have to be characterised and the protein purified to homogeneity by alternative methods such as ion exchange or reverse phase chromatography. The reason to that work on HGF-K was discontinued was not due to the heterogeneity of the sample, but rather due to the direction of the project and what experiments were planned.

4.2.3 HGF-NK-C

HGF-NK-C refers to the version of HGF-NK cloned in this project, to distinguish it from HGF-NK-G cloned by the Gherardi research group in Cambridge. HGF-NK-C contains 178 amino acids, has a molecular mass of 20515.5 Da and a theoretical pI of 9.49.

HGF-NK-C was expressed in the same manner as HGF-N, but no ^{15}N-labelled protein was produced. Initial SDS-PAGE analysis of mini-scale supernatants showed no bands of the correct size (Figure 4.18).

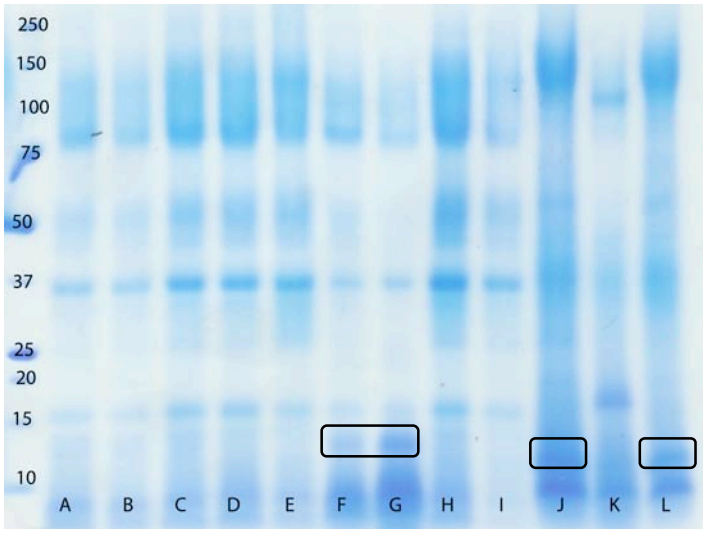


Figure 4.18 SDS-PAGE analysis of the *Pichia*-cell supernatants after mini-scale expression of HGF-NK-C. The first lane from the left is a protein ladder and numbers indicate the kDa weight of the protein. The letters for the remaining lanes are supernatants from different clones. The protein is expected to appear at 20-25 kDa, but no corresponding bands can be observed. Some new low molecular weight species can be observed in lanes F, G, J and L, as indicated by rectangles. Clone J was chosen for large scale protein production.

The SDS-PAGE analysis did however show some new low molecular weight bands for some of the clones, for example F, G, J and L in Figure 4.18, suggesting that the protein might be degraded. To investigate further, a larger shaker culture of clone J was set up to produce more protein for mass spectrometry analysis. The culture was harvested and loaded onto a 1 ml HiTrap heparin-Sepharose column and eluted with a NaCl gradient (Figure 4.19).

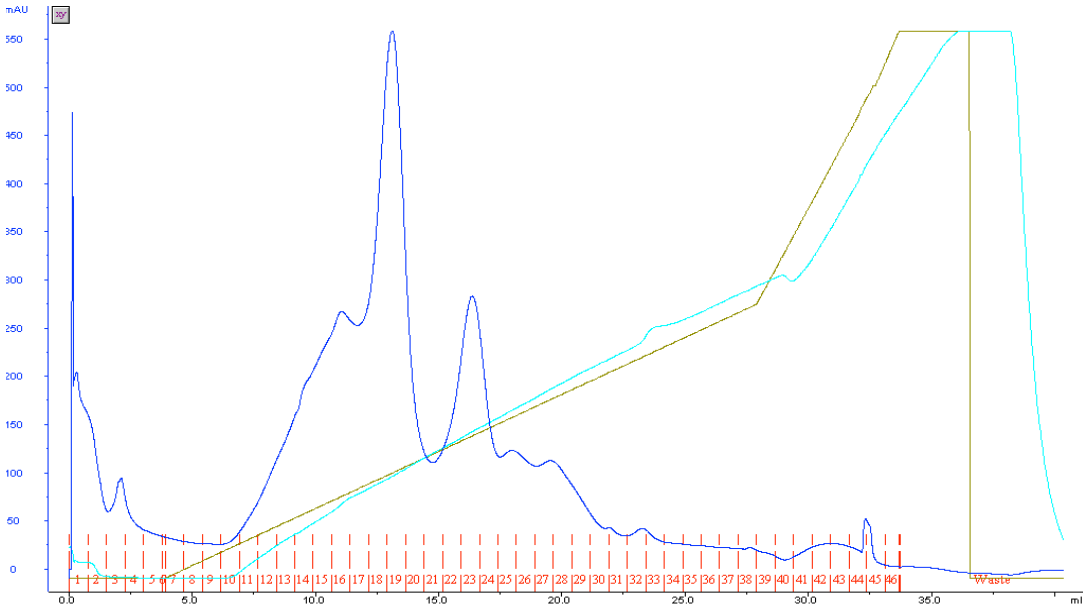


Figure 4.19 Elution of HGF-NK-C from a 1 ml HiTrap heparin-Sepharose column.

The chromatogram shows two distinct peaks and the fractions were analysed by SDS-PAGE in Figure 4.20.

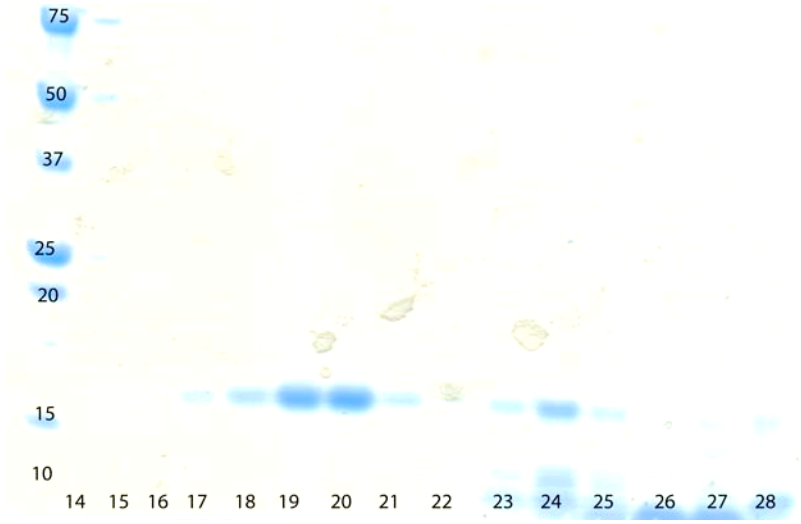


Figure 4.20 SDS-PAGE analysis of fractions 14-28 in Figure 4.19. A protein ladder with its molecular weights is in the lane furthest to the left. The remaining lanes are numbered according to the fraction number from the chromatogram in Figure 4.19 Two proteins of a similar size but of different heparin affinity can be observed.

The HGF-NK-C construct is expected to run at approximately 20 kDa, but bands can only be seen at 15 kDa for the early fractions and below 10 kDa for late fractions on the SDS-PAGE gel. It was important to clarify what these two peaks are in order to solve a possible truncation problem, may it be protein degradation or premature protein translation termination. FTICR mass spectrometry was employed on fractions 19 and 23 in Figure 4.19.

Fraction 19 had an experimental mass of 9818.68 Da and an ECD fingerprint matching the sequence:

DYIRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEHSFLPSSYRGKDLQENY
CRNPRGEEGGPWCFTSNPEVRYEVCDIPQCSE

Fraction 23 had an experimental mass of 9539.58 Da and an ECD fingerprint matching the sequence:

IRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEHSFLPSSYRGKDLQENYCR
NPRGEEGGPWCFTSNPEVRYEVCDIPQCSE

The two sequences correspond closely to the K1 domain. It was concluded that it is likely a protease degradation problem acting at two different sites in the protein sequence:

GQRKRRNTIHEFKKSAKTTLIKIDPALKIKTKKVNTADQCANRCTRNKGLPF
TCKAFVFDKARKQCLWFPFNSMSSGVKKEFGHEFDLYENK^DY^IRNCIIGK
GRSYKGTVSITKSGIKCQPWSSMIPHEHSFLPSSYRGKDLQENYCRNPRGEEG
GPWCFTSNPEVRYEVCDIPQCSE

Where the ^ indicates the cleavage sites, red is the N domain, blue is the K domain and green is in the linker between the two domains.

The cleavage sites are in the linker region between the N and the K1 domain, which according to the ExPASy database consists of the residues in green DYIRNC, but the neighbouring residues are also relatively flexible and exposed in the crystal structure, leaving the linker prone to protease cleavage. There are different ways to minimise

protein degradation in *P. pastoris* [222], [237-240] and three feasible possibilities were identified to avoid the degradation problem:

1. Screen different medium conditions that minimise protein degradation.
2. Mutate the degradation site.
3. Use a protease deficient *Pichia pastoris* strain.

As there is less work and cost in screening different media, that possibility was explored first. One aim with the HGF-NK-C construct was to perform NMR studies on ^{15}N-labelled protein, therefore ideally no nitrogen source could be added to the medium. Seven different types of yeast media or conditions were investigated:

1. BMM. As a reference medium.
2. BMM supplemented with AEBSF. AEBSF is an irreversible serine protease inhibitor like PMSF, but it is more active.
3. BMM supplemented with AEBSF and EDTA. EDTA would also inhibit the activity of metalloproteases.
4. MM, contains no phosphate buffer and the pH goes acidic, which minimises proteolysis as many proteases work better at high pH. The yeast cells can cope well with the lowered pH.
5. BMM with casein amino acids. The casein amino acids are short peptides and compete with the expressed protein for proteases. The proteases degrade the casein amino acids rather than the protein and the casein amino acids are then purified away during the protein purification steps. This approach would not allow ^{15}N labelling.
6. BMM with daily harvesting. Harvesting the culture on a daily basis would minimise the time that the expressed protein is exposed to proteases.
7. BMMY, i.e. BMM supplemented with yeast extract and peptone or tryptone. The principle is the same as for casein amino acids supplementation, the proteases are saturated with peptone/tryptone to minimise degradation of the expressed protein. This approach would not allow ^{15}N labelling.

The different media conditions were screened by expressing the protein in shakers, harvesting the cultures and loading the supernatants onto a 1 ml heparin-Sepharose column that was eluted with a NaCl gradient (Figure 4.21 - Figure 4.27).

Expressing the protein in standard BMM medium gave a chromatogram with two low affinity peaks, similar to the results in Figure 4.19 (Figure 4.21). The heparin affinity is low as only the K1 domain remains intact and it has relatively low heparin affinity compared to the N domain. From experience it is known that HGF-N elutes at roughly 30-35 minutes with the used NaCl gradient, this implies that HGF-NK-C should elute at approximately the same time and that the N domain is missing. Possibly, the N domain is completely degraded.

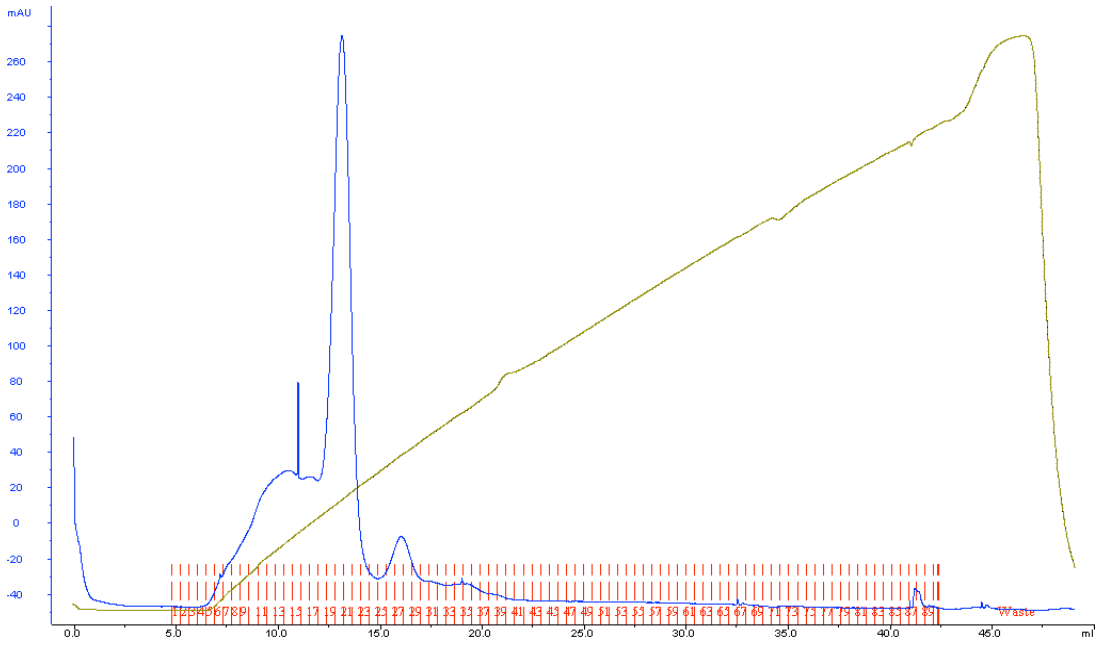


Figure 4.21 Elution of the culture grown in standard BMM from a 1 ml HiTrap heparin-Sepharose column.

To allow the protein to be expressed under minimal medium conditions, compatible with isotope labelling, the effect of AEBSF and EDTA addition to the medium were explored. The addition of these molecules to the medium might be toxic to the yeast

cells as EDTA chelates necessary trace metals and AEBSF may inhibit proteases that are involved in vital cellular processes. Comparing the chromatograms after expressing the protein in AEBSF (Figure 4.22) and AEBSF + EDTA (Figure 4.23) with that of Figure 4.21, it can be seen that expression was poor, implying that these protease inhibitors were indeed toxic. The small amount of protein that was expressed was still degraded.

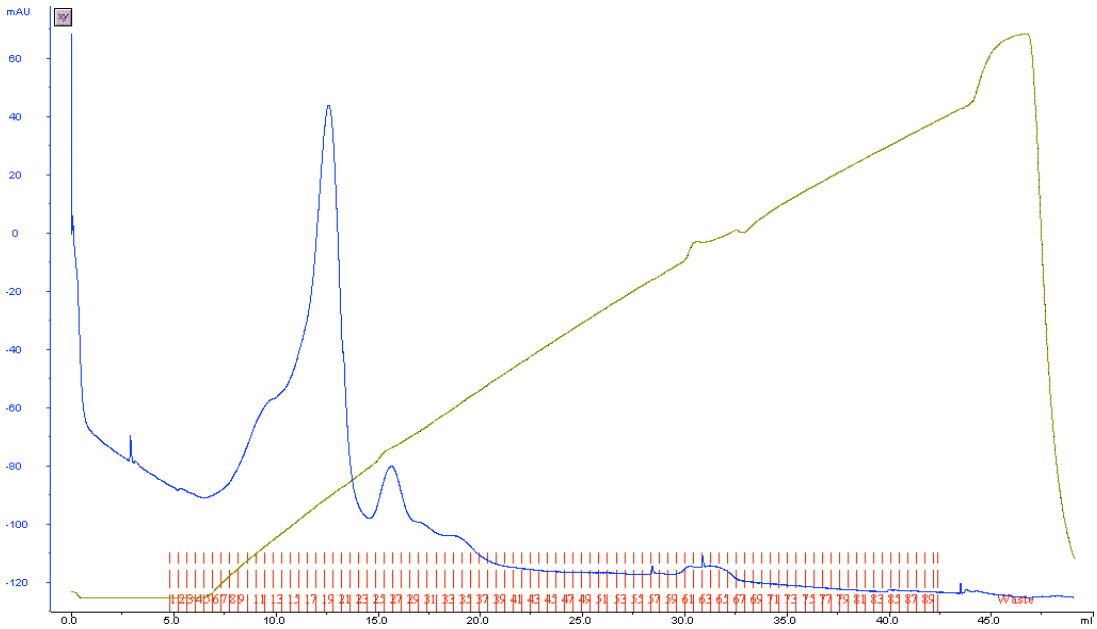


Figure 4.22 Elution of the culture grown in BMM supplemented with AEBSF from a 1 ml HiTrap heparin-Sepharose column

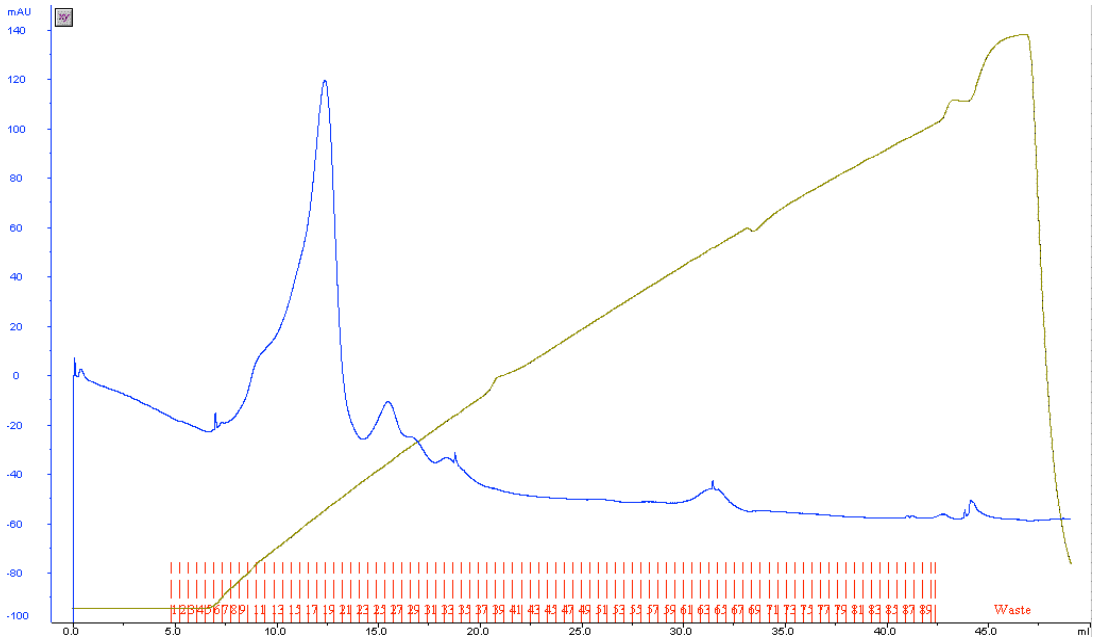


Figure 4.23 Elution of the culture grown in BMM supplemented with AEBSF and EDTA from a 1 ml HiTrap heparin-Sepharose column.

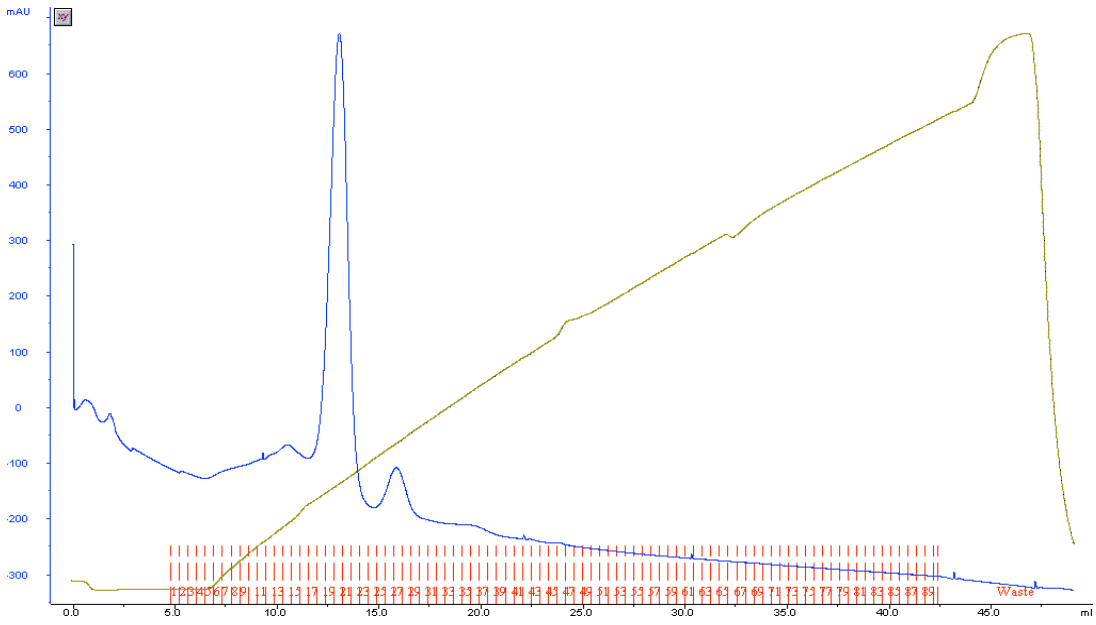


Figure 4.24 Elution of the culture grown in MM from a 1 ml HiTrap heparin-Sepharose column.

P. pastoris secretes different types of proteases that are sensitive to pH differences. By lowering the pH of the medium to 3-5, neutral and basic proteases will be less active and *P. pastoris* cells cope well with pH conditions in this range [241], [242]. With this in mind, minimal unbuffered medium (MM) was investigated, but lowering the pH did not lower proteolysis (Figure 4.24).

The Invitrogen *Pichia* manual [222] recommends adding casein amino acids to the medium to avoid proteolysis and stabilise the expressed proteins. The approach is often applied in biochemical research, for example BSA is often added to protein samples as a stabilising agent. Supplementing the medium with casein amino acids helped to minimise proteolysis and a native HGF-NK-C peak eluting at 30-35 minutes appeared in the chromatogram (Figure 4.25). As casein amino acids contain carbon and nitrogen, this approach would not allow isotope labelling of the protein and does not completely eliminate the proteolysis. The peaks in the chromatogram are very broad, implying that something was likely wrong with the chromatography run.

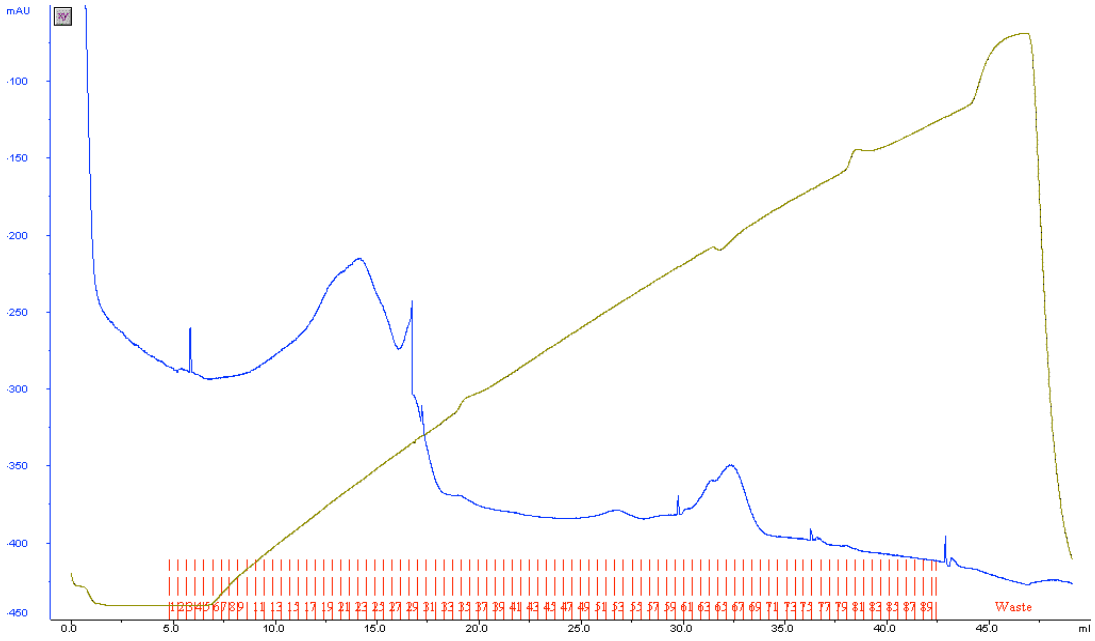


Figure 4.25 Elution of the culture grown in BMM supplemented with casein amino acids from a 1 ml HiTrap heparin-Sepharose column.

To minimise the time the expressed protein was in contact with the *Pichia* proteases, standard BMM medium was used, but the culture was harvested on a daily basis. The advantage of this approach is that it allows for isotopic labelling. The approach worked and approximately 40% of the protein was full-length native HGF-NK-C eluting at 30-35 minutes (Figure 4.26). The disadvantage of this approach is that it is stressful for the cells and labour intensive.

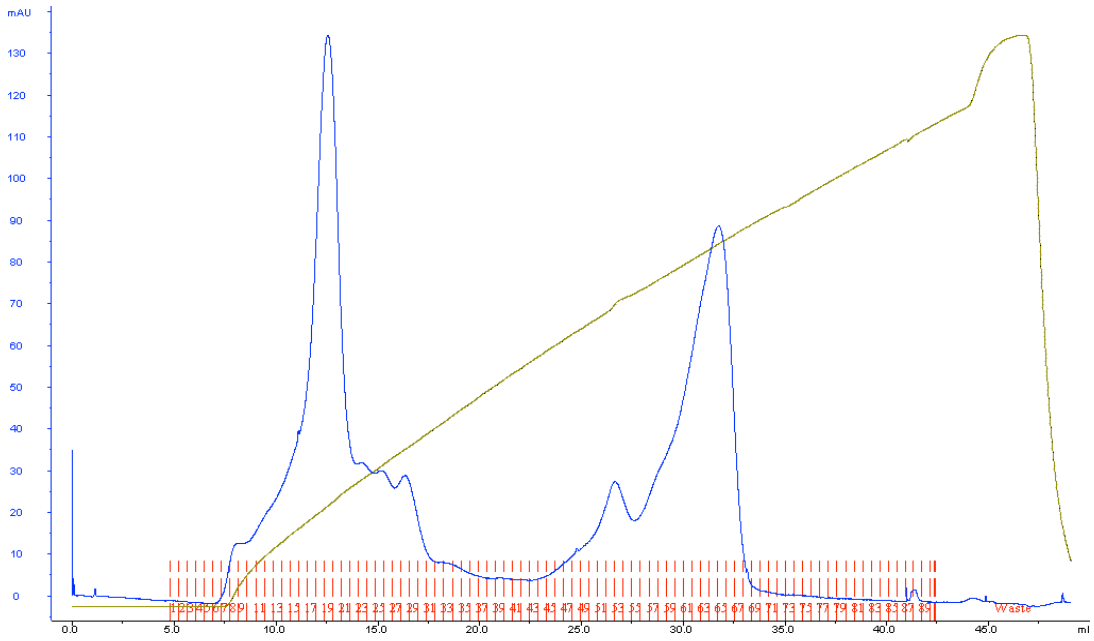


Figure 4.26 Elution of the culture grown in BMM, but harvested on a daily basis from a 1 ml HiTrap heparin-Sepharose column.

The most successful approach was to induce expression in BMMY medium, which is similar to supplementing BMM medium with casein amino acids, the proteases work on the peptone or tryptone instead of the expressed protein. The chromatogram in Figure 4.27 shows a major HGF-NK-C peak eluting at 30-35 minutes. There is also a shoulder at 27 minutes. MALDI-TOF and FT-ICR mass spectroscopy analysis showed that the peak eluting at 30-35 minutes contained the full, undegraded HGF-NK-C sequence and that the shoulder at 27 minutes contained the same heterogeneous N-terminuses as HGF-N (data not shown). This approach does not allow isotope labelling, but minimises proteolysis effectively.

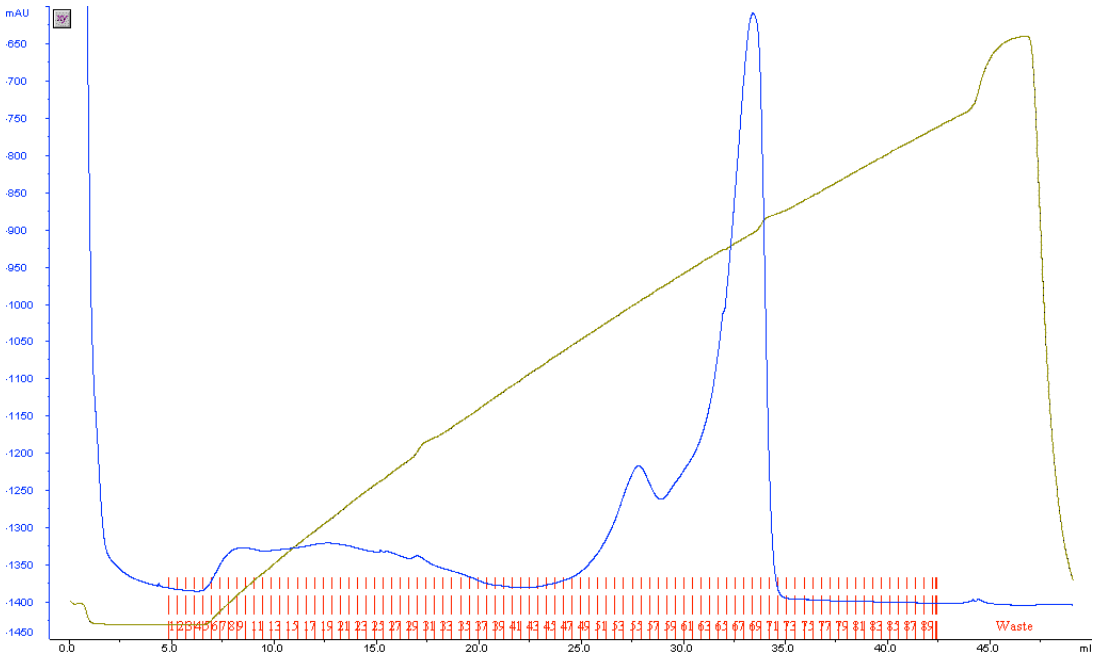


Figure 4.27 Elution of the culture grown in BMMY from a 1 ml HiTrap heparin-Sepharose column.

A fermentation of HGF-NK-C was carried out to produce larger quantities of the protein in fermentation minimal medium. The cells were grown just like for the other protein constructs, but before induction, yeast extract and peptone were added to the culture to mimic BMMY medium, which was most effective in reducing HGF-NK-C proteolysis in shaker flasks. The protein was captured on a heparin-Sepharose resin that had to be discarded afterwards due to peptone and yeast extract contamination. From the chromatogram in Figure 4.28, it can be seen that more than 50% of the protein is not native, likely having N-terminal truncations or other degradations and modifications. The reason why more degradation was seen under fermentation expression conditions compared to shaker flask protein expression conditions may be due to the higher cell densities.

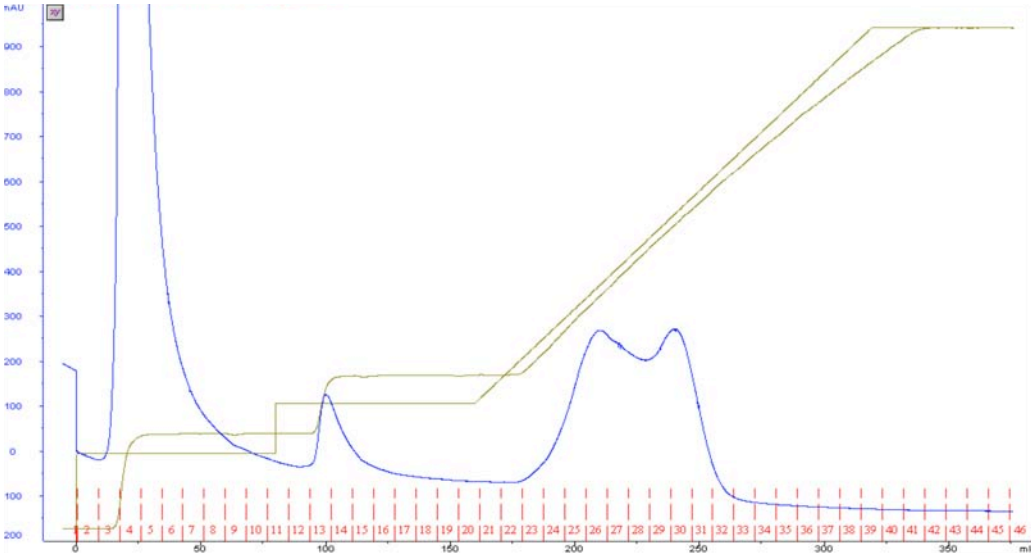


Figure 4.28 Elution of fermented HGF-NK-C supernatant from a heparin-Sepharose resin.

To further purify the protein, MonoS chromatography was carried out in a similar manner to HGF-N. The chromatograms in Figure 4.29 show similar features to HGF-N, with the later heparin purified fractions being purer than the earlier fractions.

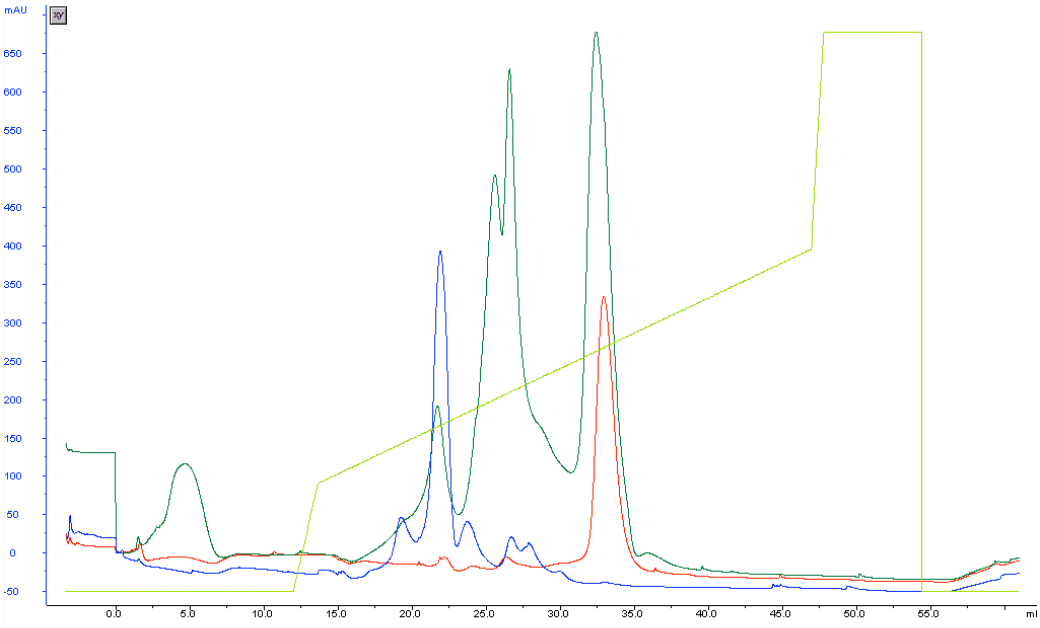


Figure 4.29 MonoS chromatography of HGF-NK-C expressed in a fermenter. The blue trace represents fractions 23-27, the green trace represents fractions 28-30 and the red trace represents fraction 31 from Figure 4.28, respectively.

In conclusion, HGF-NK-C had the same N-terminal truncation problems as HGF-N, which could not be avoided. Further, a protease site between the N and K domain was mapped and degradation at this site could be minimised to a certain extent by adding peptone or tryptone to the medium. Expression of HGF-NK-C was still relatively poor compared to HGF-N.

4.3 Characterisation of Dimeric HGF-N

During the purification of HGF-N, it was noticed that a new peak sometimes occurred during heparin affinity and gel filtration chromatography. Interestingly, the new peak displayed higher heparin affinity than HGF-N, but had a native molecular weight as determined by mass spectrometry. Initial ^{15}N-^{1}H HSQC spectra of the high affinity form showed some differences and more high affinity protein was purified to characterise the high affinity form in detail. During this purification, it was found that the high affinity form converts into the normal affinity version once isolated, but not the other way around. To confirm this observation the following experiment was performed:

1) High affinity protein was purified by heparin affinity chromatography (blue curve in Figure 4.30).
2) 50% of the high affinity protein was incubated at 37° C for 2.5 hours and 50% of the high affinity protein was incubated at 4° C for 2.5 hours.
3) Both samples were analysed by heparin affinity chromatography (Figure 4.30).

The aim of the experiments in this section is to characterise the high affinity form of HGF-N to find out what causes the extraordinary heparin affinity and how the protein can convert between a high and low affinity form. The results are important because HGF or HGF-N has previously not been shown to exist in a high and low affinity form.

It was observed that the aliquot of the high affinity protein incubated at 4° C maintained the higher affinity (black curve in Figure 4.30) while the high affinity protein incubated at 37° C converted back into the normal affinity form (red curve in Figure 4.30). The small elution volume shift for the incubated normal affinity form is probably due to less protein loading onto the column.

The conclusion from this experiment is that the protein can exist in a high affinity and a normal affinity form. The high affinity form is relatively stable at 4° C, but converts back to the normal affinity form if incubated at 37° C. It was hypothesised that the protein existed in two different conformations that interconvert slowly as has been described for some proteins [243] and in the case of HGF-N, these two different forms have different heparin affinities.

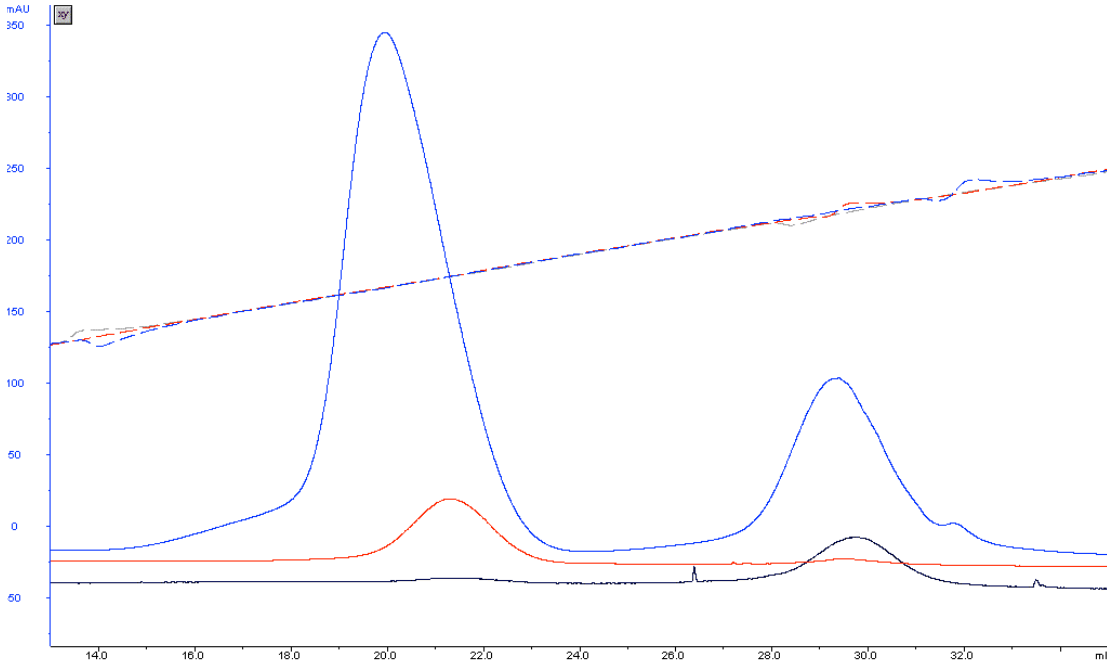


Figure 4.30 Heparin affinity chromatography of HGF-N. The blue trace is a HGF-N sample containing both high and normal heparin affinity HGF-N. The red trace is high affinity HGF-N that has been incubated at 37° C and the black trace is high affinity HGF-N that has been incubated at 4° C. The dashed curves are conductivity.

Several attempts were made to convert the low affinity form into the high affinity form. The protein was incubated on a heparin column over night at room temperature and at 4° C, it was freeze-thawed several times in high and low salt buffers, incubated with heparin DP16 at room temperature and at 4° C. Heparin affinity chromatography was also carried out in the presence of Ca^{2+} or EDTA, but none of the attempts to convert normal heparin affinity HGF-N into the high heparin affinity form were successful. It is possible that dimerisation happens during protein production and purification, or that when monomeric HGF-N is left in a liquid buffer, it slowly starts to dimerise until equilibrium is reached.

To further characterise the high and low affinity forms, gel filtration of both forms was carried out (Figure 4.31). The high and low affinity protein forms from the heparin affinity chromatography were kept cold at all times as they were spin concentrated and then injected on a gel filtration column in PBS kept at 4 °C.

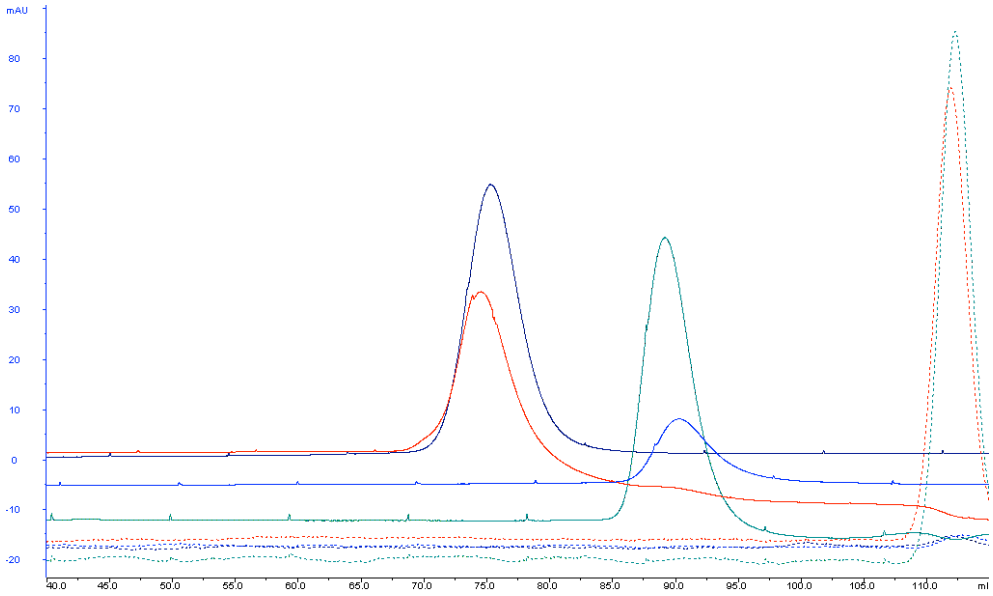


Figure 4.31 Gel filtration of different forms of HGF-N in PBS. The green curve is the normal affinity form isolated from heparin chromatography. The red curve is the high affinity form isolated from heparin chromatography. The blue curve is the fractions from the red curve after 2.5 hours at 37 °C. The dark blue curve is HGF-NK-G as a 20.5 kDa MW standard. The dashed curves are conductivity and work as internal controls.

The high affinity form eluted at the same time as HGF-NK-G, indicating that the high affinity form is dimeric since the molecular weight of HGF-NK-G is approximately twice that of HGF-N. Incubating the dimeric form at 37° C for 2.5 hours converts it into a monomeric form. It can be concluded that the high affinity form of HGF-N is a dimer.

To further characterise the HGF-N dimer, it was studied by NMR: ^{15}N labelled dimeric HGF-N was purified by heparin affinity chromatography and buffer exchanged into NMR buffer at 4 °C. As the sample was put in the NMR instrument, the temperature was maintained at 30° C and the dissociation of the dimer could be monitored by ^{1}H-^{15}N HSQC spectra (Figure 4.32). Initially, SOFAST ^{1}H-^{15}N HSQC spectra were collected to be able to see any fast changes, but the protein concentration was too low and spectral quality relatively poor. However, the dissociation process was slow enough at 30° C to be monitored by acquiring standard ^{1}H-^{15}N HSQC spectra.

Figure 4.32 shows an overlay of the initial spectrum (approximately one hour acquisition time) and the fifth spectrum acquired after 5 hours at 30° C. The latter spectrum contains peaks from both monomeric and dimeric forms of HGF-N. The overlay of the spectra indicates a slow exchange process where the crosspeaks can't be traced by gradual chemical shift differences, they just change in intensity. The off-rate of the dimer must be very slow as the dissociation can be monitored by NMR, gel filtration and heparin affinity chromatography.

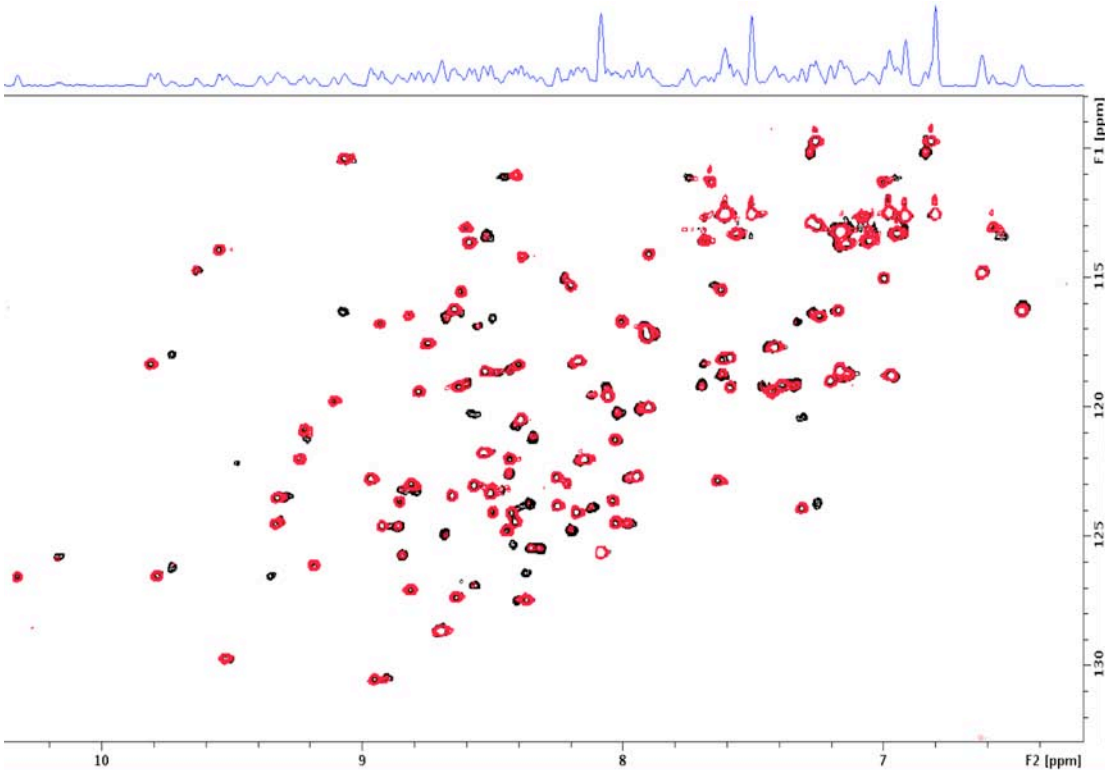


Figure 4.32 Overlaid ^{1}H-^{15}N HSQC spectra of HGF-N prepared at 4 °C. The red and black spectrum is the first and fifth of a series of 57 min spectra, respectively.

Studying the HSQC spectra, some of the new slow exchange cross peaks can be assigned with reasonable confidence, but for some new crosspeaks the spectrum is too crowded and the chemical shift differences too big, giving an unsecure and incomplete identification of affected residues. However, two consistent patches can be assigned with relatively high confidence, namely residues 63 to 68 and 86 to 90, belonging to the α2-helix/β2-strand and β3-strand, respectively. These two patches are relatively close to each other in the protein structure, but residues 86 to 90 are not surface exposed as can be seen in Figure 4.33.

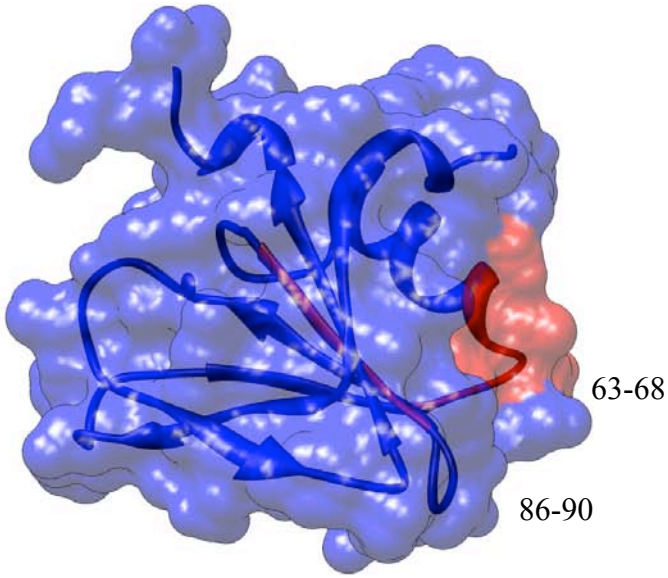


Figure 4.33 The 1BHT pdb file was rendered in UCSF Chimera 1.4 to show residues 63 to 68 and 86 to 90 in red. These are the two patches that displayed large chemical shift differences in the dimeric vs monomeric state and that could be assigned with relatively high confidence. The patch consisting of residues 63-68 is surface exposed, but the patch consisting of residues 86-90 is not and therefore has to be affected by secondary effects in the dimeric state.

The lack of a complete assignment of the affected residues prevented identification of a possible dimer interface. It was noted that the intensity of the signals increased with time and the signals became narrower (Figure 4.34). This implies that the protein tumbles faster, which supports the hypothesis that the protein goes from a dimeric state to a monomeric state. The intensity increase could also be due to chemical exchange, but this is unlikely. The black vertical line in Figure 4.34 is over a crosspeak, which decreases in intensity with time. This peak is one of the dimer peaks that changes chemical shift in the monomeric state.

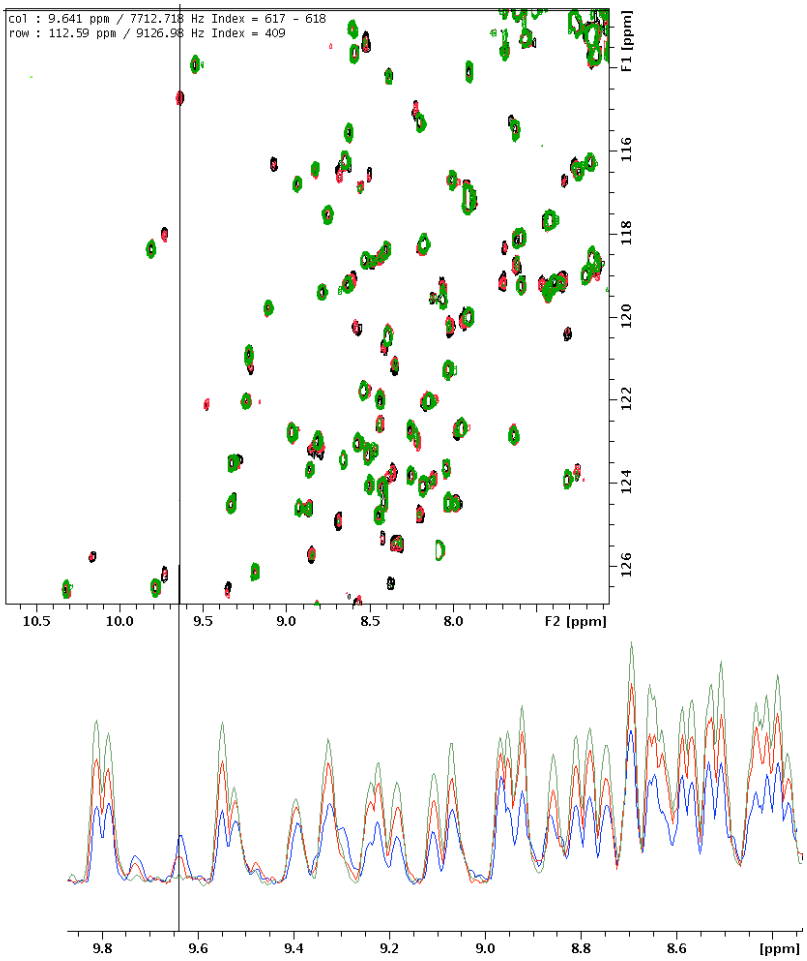


Figure 4.34 Upper: Overlaid ^{1}H-^{15}N HSQC spectra. The black spectrum is the first HSQC, the red is the third and the green is the sixth. Lower: ^{1}H projection of the HSQC spectra. The blue spectrum is from the first HSQC, the red is from the third and the green is from the sixth.

An alternative approach to mapping the HGF-N dimer interface could be to use chemical crosslinking, trypsin digestion and mass spectrometry [244]. The problem when using this approach to map a homodimer interaction is that intra and inter - protein crosslinks need to be distinguished. One way of doing that is to mix ^{14}N and ^{15}N labelled protein and only take into consideration crosslinked peptides that contain ^{14}N+^{15}N fragments and to ignore peptides composed of ^{14}N+^{14}N or ^{15}N+^{15}N fragments.

The NK1 crystal structures 1GMO, 1GMN and 1BHT PDB files were inspected for HGF-N protein-protein interactions. In the HGF-NK-G crystal structure without heparin (1BHT), a N-domain protein-protein interaction interface could be identified and UCSF Chimera 1.4 was used to search for possible hydrogen bonds in the interface (Figure 4.35). The dimer interface consists of two charged patches, 36RNTIH40 and 122KDYIRN127, that interacts with the same residues on the opposite protein molecule. These two patches do not match with the patches mapped in Figure 4.33. In crystal structures, unspecific protein-protein interactions are very common and HGF-NK1 crystallises as a head to tail dimer, where the N and K1 domains of the two protomers interact with each other.

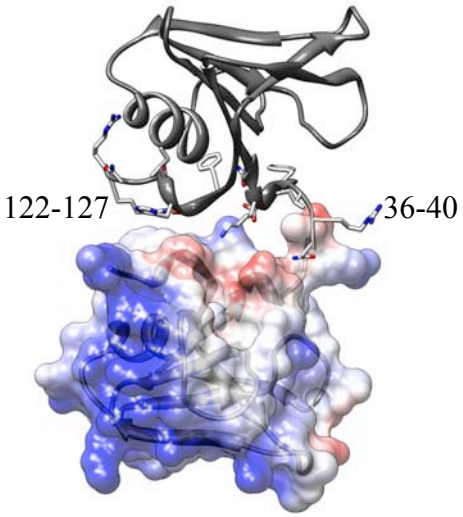


Figure 4.35 Columbic surface colouring of one protomer of the 1BHT PDB file and ribbon structure of the other protomer. Rendered in UCSF Chimera 1.4.

4.4 Purification of GAG derived oligosaccharides

HGF-N precipitates in the presence of crude heparin [21]. In order to avoid precipitation and study the HGF-N - GAG interaction in solution, oligosaccharides have to be produced. Preparing oligosaccharides by enzymatic digests also have other benefits discussed in section 1.3.8. To be able to compare the binding between different GAGs, CSA, CSC, DS and heparin were enzymatically digested and

oligosaccharides size-fractionated. In some instances, fully sulfated species from a specific size-fraction were purified from the mixture of different sulfation patterns.

A wide range of oligosaccharide lengths were sought and the digest was stopped before the GAG polysaccharides were completely digested into disaccharides. To stop the digest at the right point, aliquots of the digest were analysed on a GE Healthcare Superdex Peptide column at different time points. As can be seen in Figure 4.36, when the digest has just started, most of oligosaccharides are undigested as evident from a large broad peak eluting between 12 to 20 minutes. The disaccharide fraction can be seen eluting at 30 minutes, the tetrasaccharide at 26 minutes and the hexasaccharide at 24 minutes. Some low molecular weight contaminants can also be seen eluting at 35 minutes.

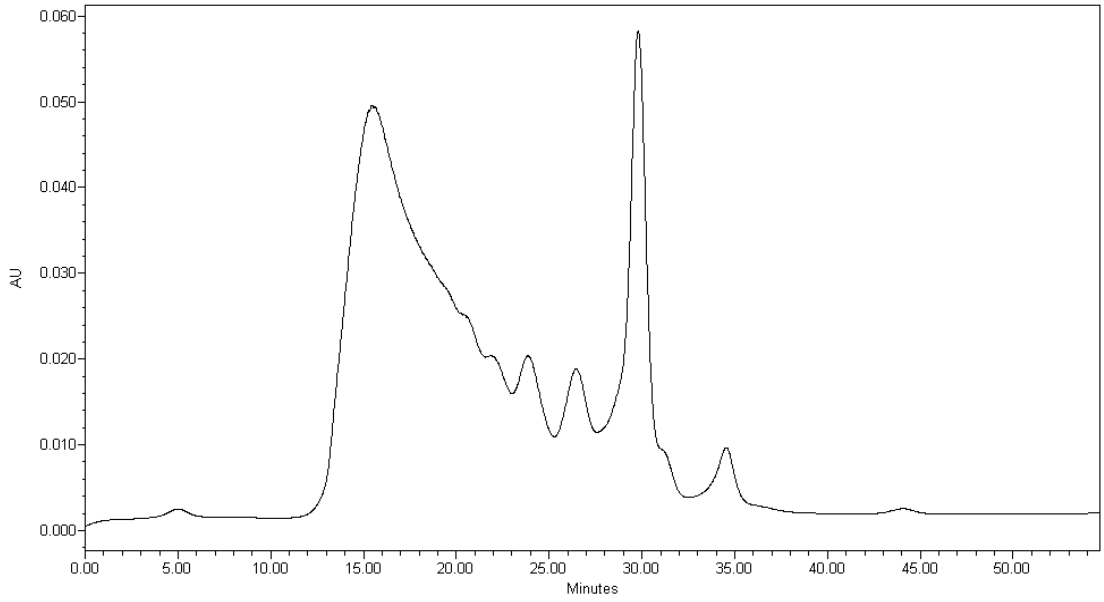


Figure 4.36 Superdex Peptide gel filtration chromatogram of a heparinase I digest of heparin one hour after starting the digest.

Comparing the 1-hour digest (Figure 4.36) with overnight digestion (Figure 4.37), it is evident that the overall absorbance has increased 10 fold. This is due to that the enzymatic cleavage produces a conjugated double bond at the non-reducing end

monosaccharide that absorbs at 232 nM and more of these conjugated double bonds have been formed. It is also clear that most of the longer oligosaccharides have been digested into shorter fragments as the broad peak eluting at 15 minutes is now gone.

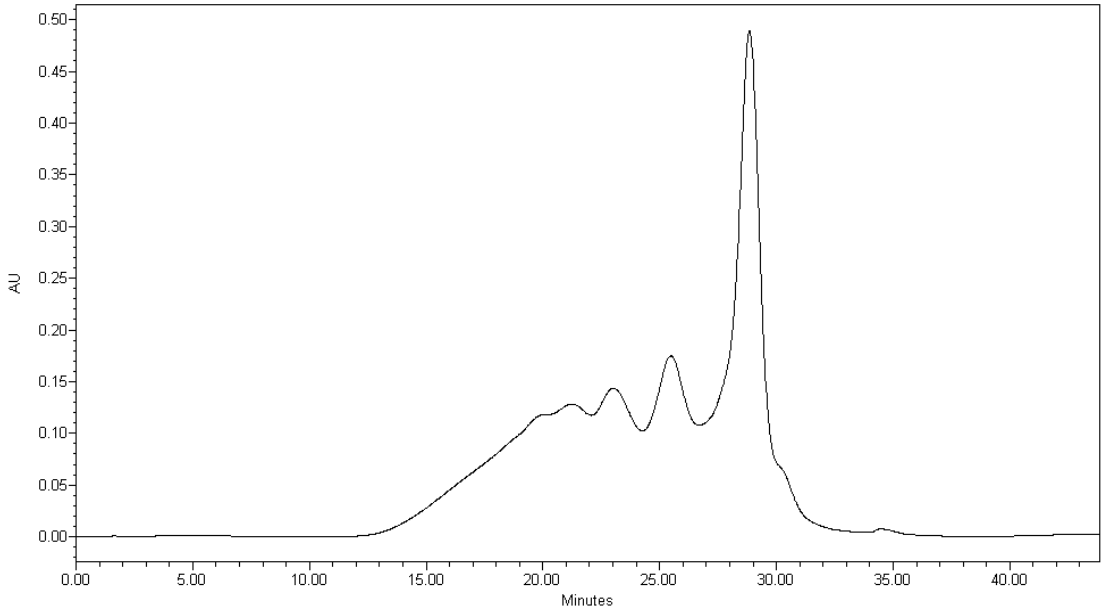


Figure 4.37 Superdex Peptide gel filtration chromatogram of a heparinase I digest of heparin after the digest has been incubated over night.

After the digest was stopped, the oligosaccharides were freeze-dried, redissolved in a small volume of H_2O and applied onto a preparative gel filtration column. Using a relatively long 180 cm BioRad P10 column, DP12 can be purified to homogeneity and DP18 can be prepared with good purity (Figure 4.38). There are small shoulders between the DP2 and DP4 fraction and the DP4 and DP6 fractions. These are likely some uncommon tri and pentasaccharides. Their appearance is hard to explain, but the presence of these oligosaccharides in CS and DS digests is well established and has been characterised by NMR [245], [246].

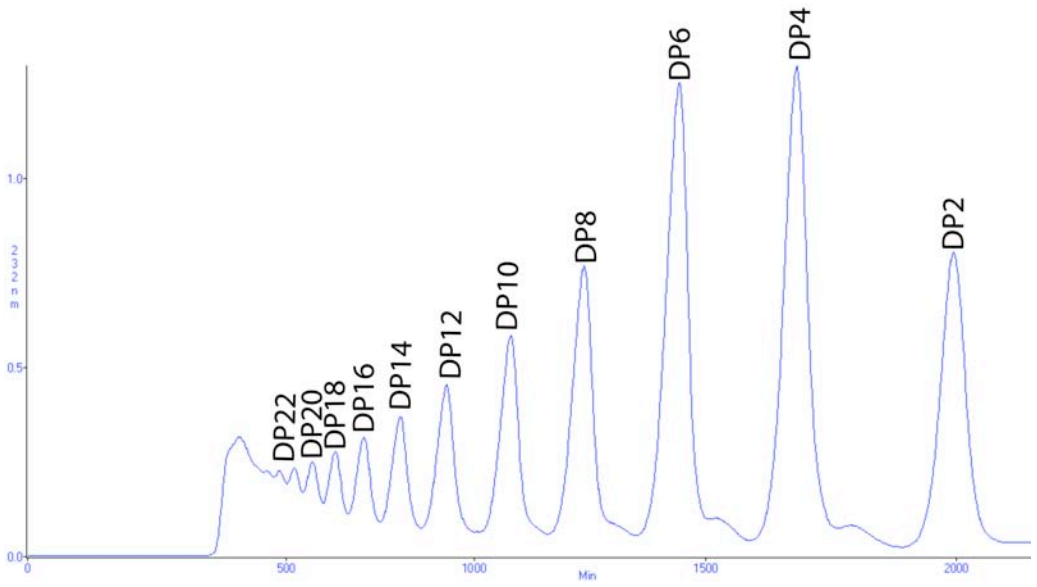


Figure 4.38 Preparative gel filtration of DS digested by chondroitinase B.

Gel filtration can separate species by size, but the sulfation pattern and carbohydrate backbone structure will still vary within each size fraction, especially for heparin and HS. These species with different sulfation patterns can be separated by strong anion exchange chromatography. Figure 4.39 shows a SAX chromatogram of a heparin DP4 fraction, separating the different species depending on the number of sulfations and backbone sugar units.

The structures of the three major tetrasaccharides separated by SAX in Figure 4.39 have been determined by NMR [247], [248]. Structures A, B and C (Figure 4.40) correspond to peaks A, B and C in the chromatogram of Figure 4.39. Tetrasaccharide A has one unsulfated GlcA sugar unit instead of sulfated IdoA, tetrasaccharide B lacks the C6 sulfate at the reducing monosaccharide and tetrasaccharide C is fully sulfated.

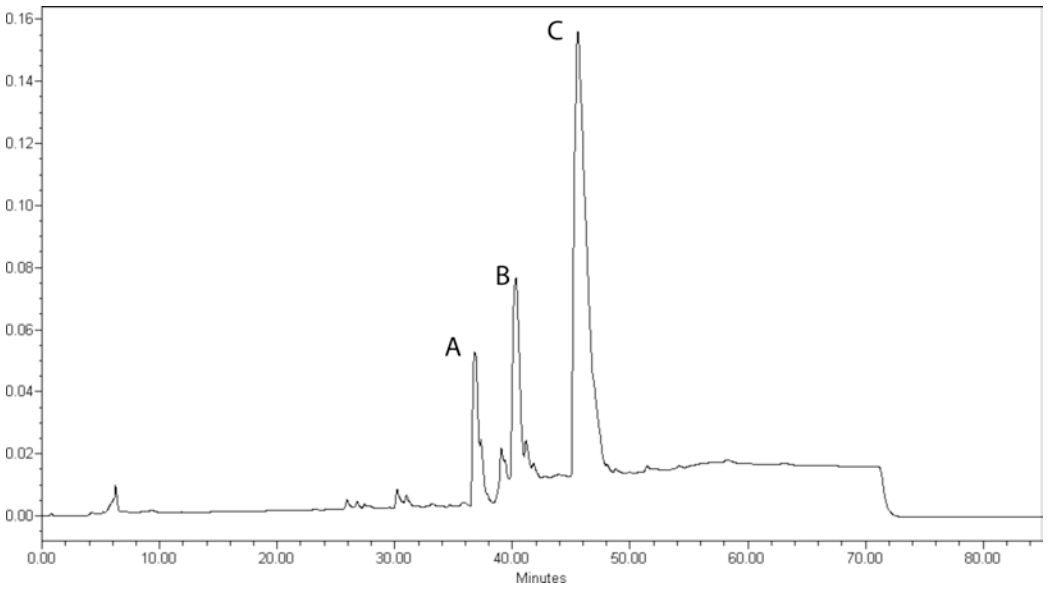
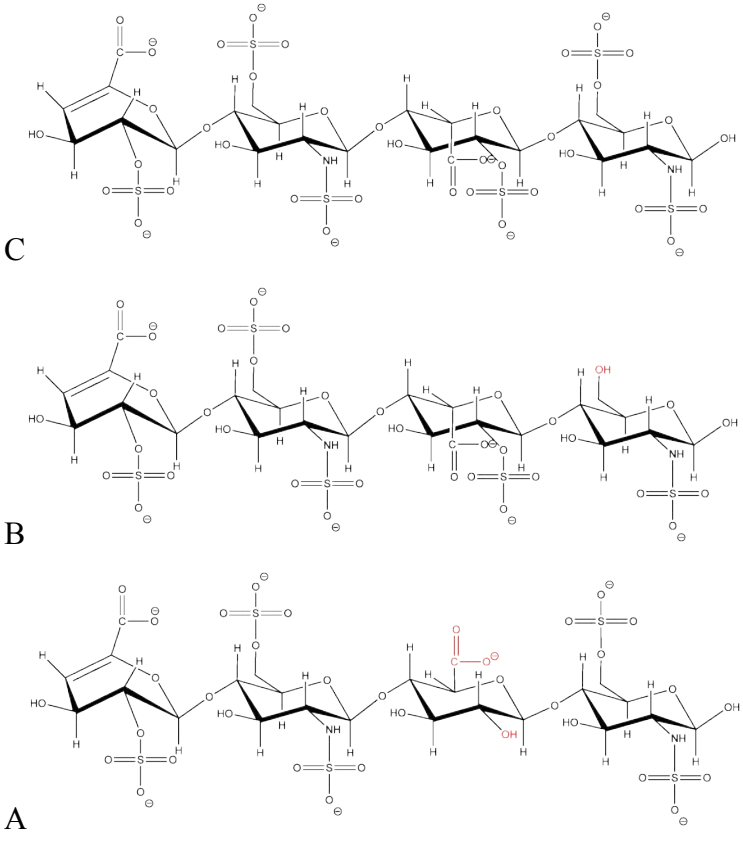


Figure 4.39 SAX chromatogram of the heparin DP4 fraction obtained by gel filtration.



C

B

A

Figure 4.40 Structures of heparin-derived tetrasaccharides separated by SAX in Figure 4.39.

The homogeneity and purity of the enzymatically derived GAGs was assessed by SAX chromatography as show above, MALDI-TOF spectrometry (Figure 4.56) and by ^{1}H NMR (Figure 4.41). The latter technique is particularly suitable to assess the overall purity of samples and to identify any non-carbohydrate contaminants as all hydrogen containing molecules can be seen, whereas MALDI-TOF and HPLC are more selective techniques and impurities can go undetected. The enzymatically derived oligosaccharides were generally very pure according to MALDI-TOF and HPLC analysis. Some small molecule contaminants could often be seen in ^{1}H spectra, resonating between 2.4 and 3 ppm (Figure 4.41).

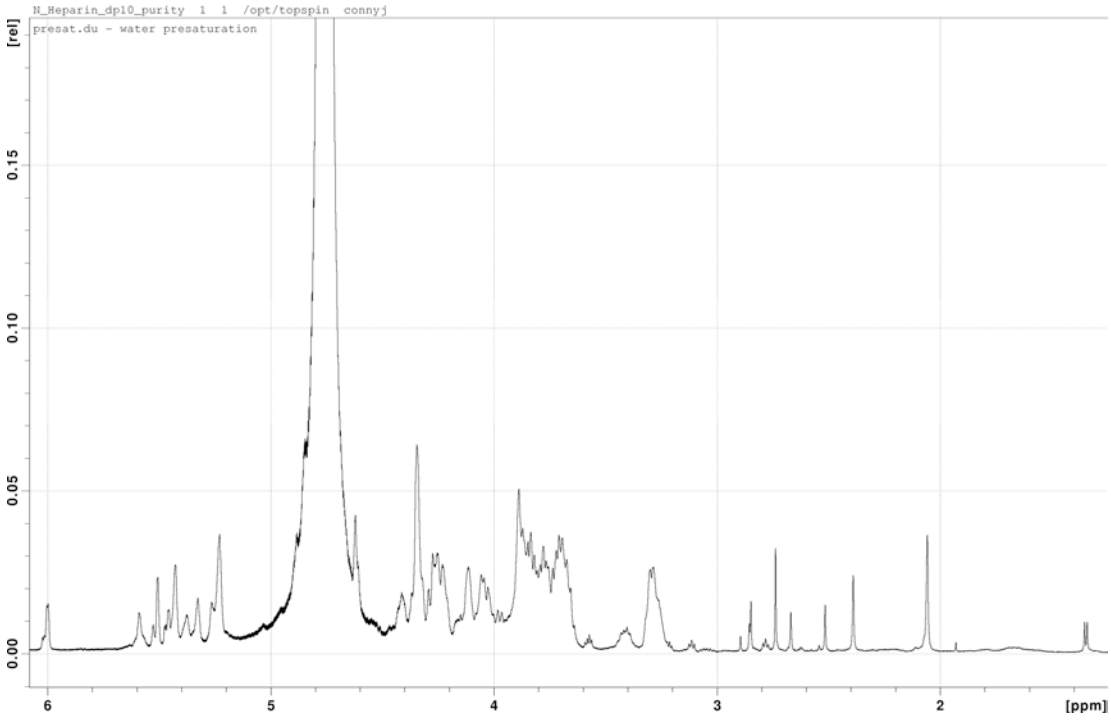


Figure 4.41 ^{1}H NMR spectrum of enzymatically derived heparin DP10.

The complexity of the different species in a specific size fraction increases quickly with oligosaccharide length. The DP4 in Figure 4.39 contains three major peaks, the DP6 in Figure 4.42 contains eight major peaks and the DP10 in Figure 4.43 contains a forest of undistinguishable peaks. To make the highly charged DP10 species elute

from the column, a stronger salt gradient had to be used compared to the DP4 and DP6 chromatograms.

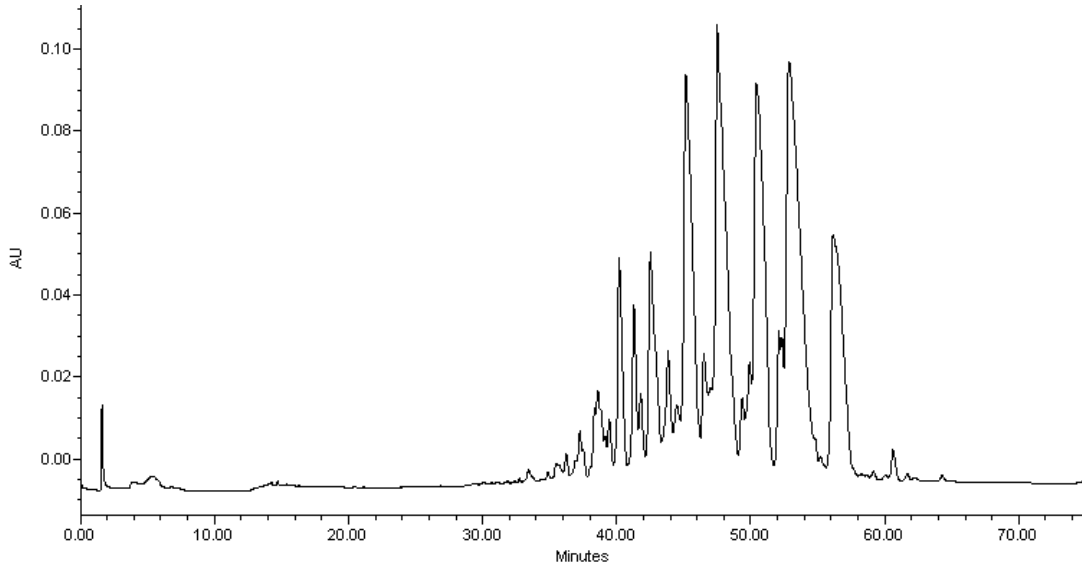


Figure 4.42 SAX chromatogram of heparin-derived hexasaccharide obtained by gel filtration.

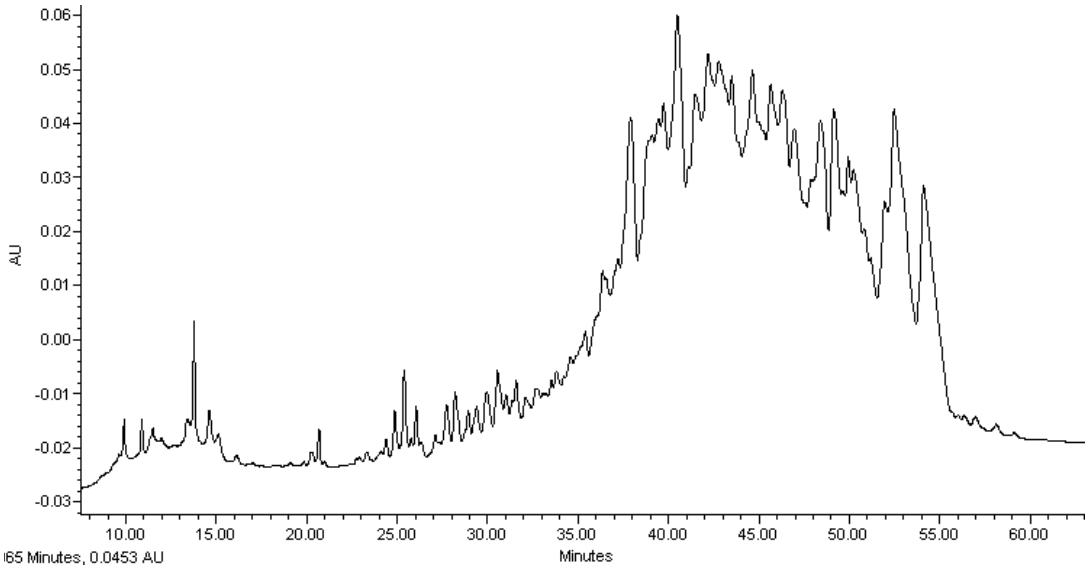


Figure 4.43 SAX chromatogram of heparin-derived decasaccharide obtained by gel filtration.

CS and DS are much more homogeneous GAGs than heparin and HS. They are also less sulfated. This results in less variation of oligosaccharides produced by enzymatic digestion as can be seen in the SAX chromatograms of the decasaccharide fractions shown in Figure 4.44.

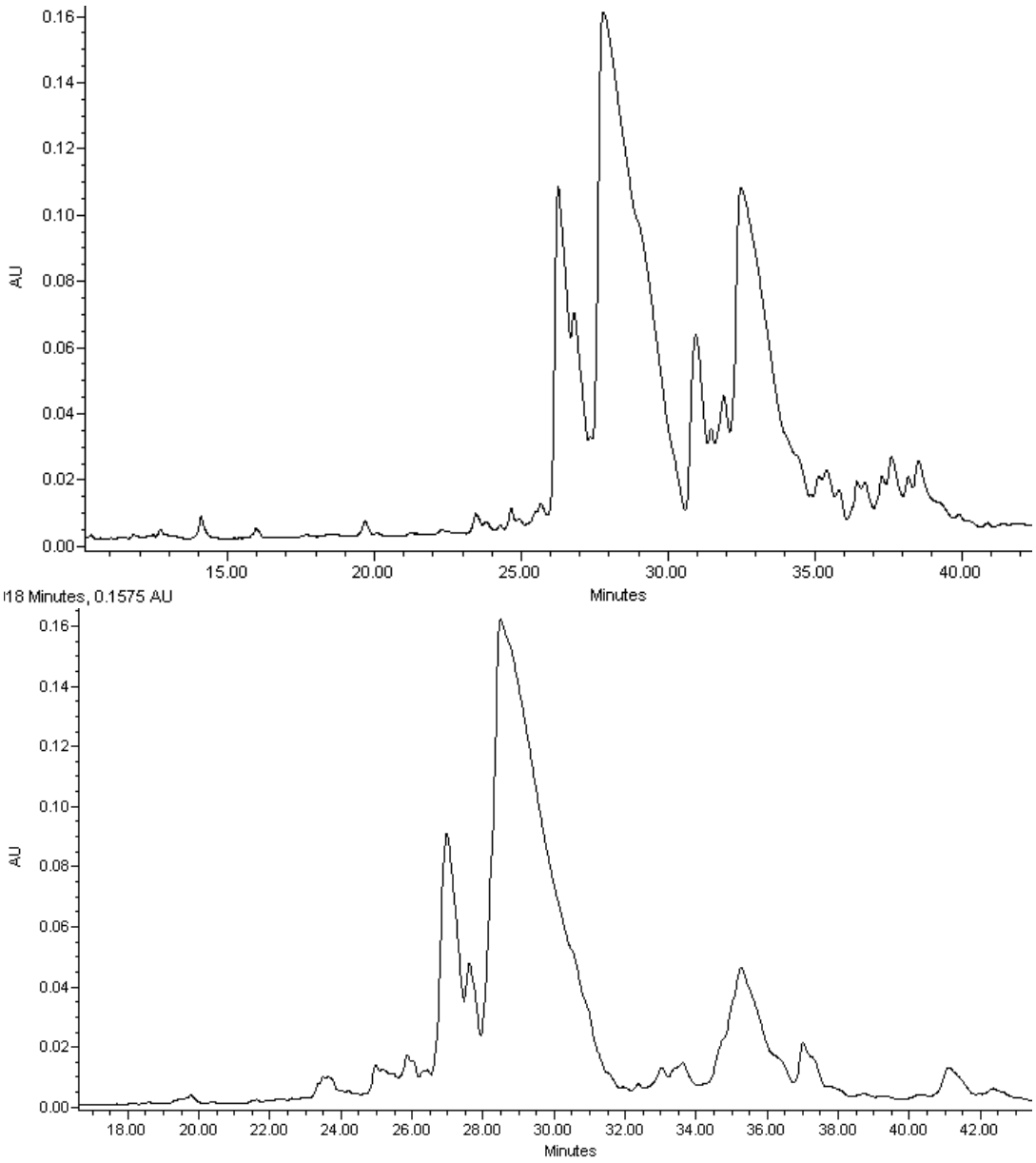


Figure 4.44 SAX chromatogram of CSC (upper) and DS (lower) decasaccharides obtained by enzymatical digestion and gel filtration.

4.5 Affinity Chromatography

4.5.1 In-house Prepared DP16 GAG Affinity Columns

There are many ways to compare GAG/heparin affinity between proteins and one of the most basic and straightforward ways is affinity chromatography. Heparin affinity chromatography has been used to study the HGF-GAG interaction in many previous studies, but no study has used the method to objectively compare the affinity between several different GAGs and HGF protein constructs. The aim of these experiments was therefore to produce GAG affinity columns with oligosaccharides of similar length so that the HGF-N and HGF-NK-G affinity for different GAGs could be quantitatively compared. The results are important in that it is the first time that a wide range of GAG affinity columns are used to compare the HGF affinity for different GAGs. Further, the results clarify the difference in GAG affinity for the different protein constructs HGF-N, HGF-NK-C and HGF-NK-G.

Heparin affinity chromatography typically relies on heparin chains immobilised to Sepharose or a similar type of bead. The studied protein binds to the heparin chain and is eluted with an increasing concentration of NaCl. The concentration required for elution is generally proportional to the K_D of the interaction. High affinity interactions with K_D values in the range of 0.1 μM to 1 nM typically require at least 1 M NaCl to elute the bound protein from the column [3]. Proteins with low heparin affinity (K_Ds in the range of 100 μM to 10 μM) either do not bind at physiological salt concentrations or require less than 0.5 M NaCl to elute. The method makes the assumption that protein - heparin interactions are entirely ionic, which is not completely true as evident from the Fondaparinux/Idraparinux example in section 1.3.7.

Crude heparin and different GAG DP16 oligosaccharides were biotinylated on the reducing end and loaded onto streptavidin columns to prepare GAG affinity columns. This allowed easy affinity comparison between different GAGs and protein constructs. Table 4.3 summarises the different columns that were used.

Table 4.3 The different GAG affinity columns and their corresponding colour in the chromatograms.

Standard heparin column from GE Healthcare	Dark blue
Heparin-biotin-streptavidin linked via free amines	Pink (only HGF-N)
Heparin DP16-biotin-streptavidin	Red
Chondroitin sulfate C DP16-biotin-streptavidin	Green
Chondroitin sulfate A DP16-biotin-streptavidin	Grey
Dermatan sulfate DP16-biotin-streptavidin	Orange

The decahexasaccharide fraction was chosen for all GAGs for affinity chromatography, AUC and ITC experiments to make comparison between the different experiments easier. As heparin, chondroitin and dermatan -sulfate chains vary widely in length, using only one defined length makes comparison between the different GAGs easier and more objective. Furthermore, DP16 is the longest heparin oligosaccharide that can be prepared with reasonable purity in our laboratory (Figure 4.38). A relatively long GAG chain was preferred over shorter oligosaccharides to mimic *in vivo* conditions. Using a long oligosaccharide also ensures that the whole GAG binding site of the protein will be spanned and that highly sulfated GAG motifs will be preserved.

There are many ways to immobilise GAGs for affinity chromatography, SPR and similar experiments [249]. The most straightforward way to introduce only one linkage to the matrix is by reductive amination at the reducing end carbonyl. By using NHS-chemistry, utilising the odd free amine that appears in heparin chains, a limited number of linkages are created. The GAG can either be directly linked to the matrix or surface, or a biotin with a linker can be conjugated to the GAG, subsequently utilising the very tight biotin-streptavidin interaction to immobilise the

GAG to the matrix [250]. In these experiments, the biotin conjugation approach was selected as the reaction was relatively straightforward and the biotinylated GAGs could potentially be used for other experiments such as SPR if required. Reductive amination of GAGs have been used to attach biotin affinity labels and other types of labels such as reporter labels to allow many different types of experiments [168], [251-262]. Reductive amination is described in detail in section 4.11.1. A schematic explanation of the process and the structures of the biotin labels used in this experiment are shown in Figure 4.45.

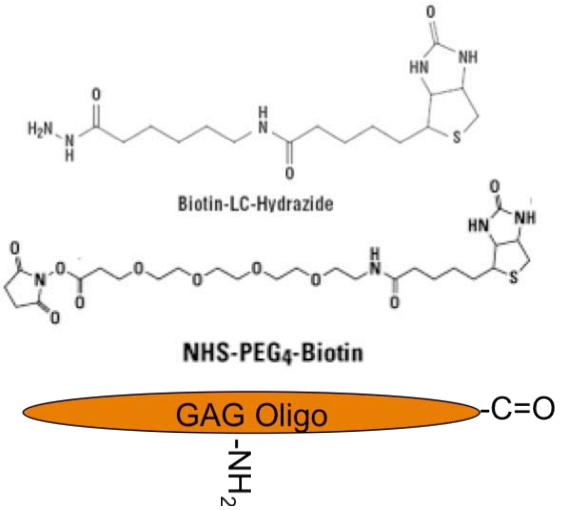


Figure 4.45 Structure of the biotin labels and targeted reactive groups of GAG oligosaccharides. Biotin-LC-Hydrazide was used for conjugating biotin to the reducing end and NHS-PEG4-Biotin was used for biotinylating heparin on free amines.

The reducing end conjugation has the advantage of producing less steric hindrance compared to the free amine conjugation as it leaves most of the GAG chain free for protein interaction. The disadvantage of the reducing end conjugation is that it opens the reducing end sugar unit. This could be a problem for shorter oligosaccharides, but modifying one out of 16 monosaccharide units is unlikely to pose a problem. Another potential problem is that the conjugation reaction conditions are relatively harsh compared to the NHS-chemistry conditions. Although some researchers report successful biotinylation under mild reductive amination conditions [11], [263], [264],

sufficient biotinylation under such conditions was not achieved with a CSA DP4 model system in this study. Possibly, the reaction could have been incubated longer and the pH adjusted to achieve better conjugation under milder conditions.

To confirm that the reaction conditions described in section 3.4 did not alter the GAG structure, a heparin hexasaccharide was exposed to the same reaction conditions, but without the biotin label (biotinamidohexanoic acid hydrazide). The SAX chromatograms and MALDI-TOF spectra of the GAG were identical before and after the reaction, indicating that the reaction conditions did not alter the structure (data not shown).

All proteins clearly bound to the prepared GAG affinity columns and eluted as sharp symmetrical peaks, except when heparin was biotinylated via free amines, as seen in Figure 4.46. HGF-N and HGF-NK-C elutes from the standard GE Healthcare heparin column at approximately 1 M NaCl, but from the prepared GAG columns at approximately 750 mM NaCl. All proteins bound tighter to the standard heparin column from GE Healthcare than to the prepared heparin DP16 column. This can be explained by:

A) On the GE Healthcare column, the heparin chains are longer and can contain more complete high affinity sites from both a quantity and quality point of view.

B) Differences in heparin batches and sources: it is well known that there is a difference in structural features between different heparin sources and batches [170], [171], [156].

C) The heparin is linked to the resin in a different manner.

D) As the heparin on the prepared column is shorter, conjugated to a biotin molecule and linked to streptavidin there is a larger possibility that the streptavidin and resin generate steric hindrance.

E) A difference in the abundance and quantity of immobilised heparin chains.

HGF-N and HGF-NK-C elute at the same time from the standard heparin column, which is expected as the main heparin binding site is thought to be in the N domain of HGF [11], [21], [14], [42], [265]. HGF-NK-C elutes somewhat later on the in-house prepared columns compared to HGF-N, but at the same time on the standard GE healthcare column. This could be explained by more sterical hindrance on the in-house prepared column or by experimental variation.

In Figure 4.46 it can clearly be seen that HGF-N and HGF-NK-C has higher affinity for all GAGs compared to HGF-NK-G. This is likely due to the difference in the N-terminal residues between the different constructs as discussed in section 4.5.2.

In following the binding of individual proteins to different GAG columns in Figure 4.46, there is a general trend that all protein constructs bind tightest to the standard GE Healthcare heparin column followed by the in-house prepared heparin column. All proteins bind with similar affinities to CSA, CSC and DS.

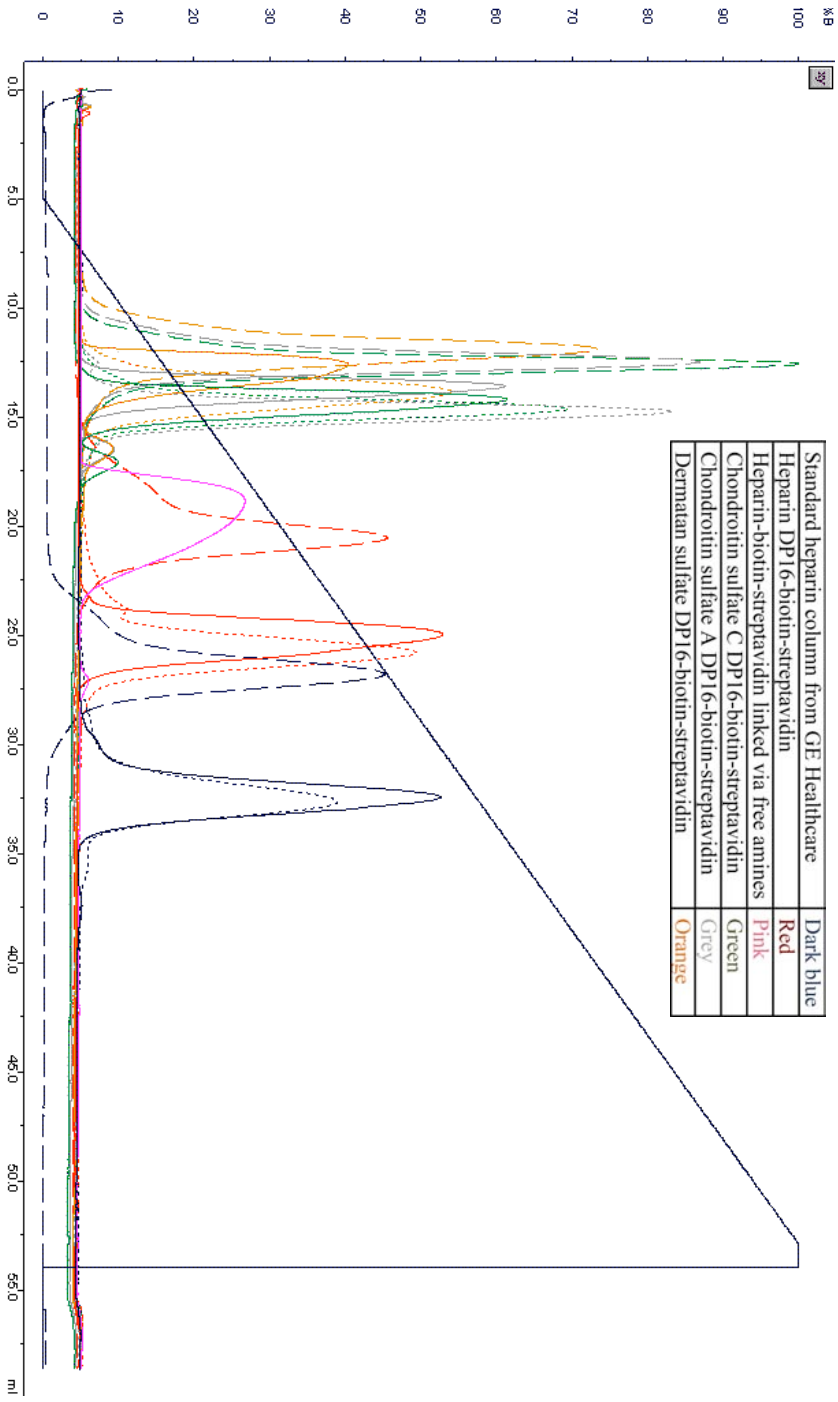


Figure 4.46 Chromatograms showing HGF-N (solid line), HGF-NK-C (dotted line) and HGF-NK-G (dashed line) on all the different GAG affinity columns. The elution volume is shown on the X-axis and the UV response on the Y-axis. The gradient going from buffer A to buffer B is shown as a solid blue line. A general trend is that all proteins bind with the highest affinity to heparin and with lower but similar affinities to CSA, CSC and DS. HGF-NK-G binds with lower affinity to all GAGs compared to HGF-NK-C and HGF-N.

Turning our attention to the HGF-N chromatogram for the column where biotin was linked to free amines of heparin (pink trace in Figure 4.46), it is clear that the binding is impeded compared to conjugating the biotin to the reducing end even though crude heparin was used instead of DP16. This could possibly be explained by the fact that each heparin polymer can contain several free amines spaced along the whole length of the chain. Immobilisation could constrain the heparin chain, forcing it to lie in a more horizontal mode along the matrix, creating steric hindrance for protein binding with the biotin linker, streptavidin and the matrix.

There are small, but noticeable differences in the binding of HGF-N to chondroitin A and C: the protein seems to bind tighter to the latter column. Whether this is due to experimental variation or if it reflects real differences in the binding is difficult to say. One could possibly argue that the 6-sulfate of chondroitin C is more amenable to interact with the protein since it is attached to a CH_2 group and this additional flexibility could strengthen the binding. On the other hand, the affinities for both chondroitin A and C are the same for HGF-NK-C, suggesting that the difference for HGF-N might be a small experimental variation.

There are very few, if any reports at all on that HGF binds to CS. Some studies suggest that HGF only binds to HS like GAGs [266], whereas other studies suggest that HGF binds to DS [12]. The main difference between CS and DS is that CS contains glucuronic acid backbone sugar units that are less flexible compared to the iduronic acids sugar units in DS. It has been argued that this allows DS to adopt a more favourable conformation for binding with higher affinity to HGF. However, in this study it was found that HGF binds with similar affinity to CS and DS. It is noticeable that the dermatan peaks are somewhat wider than the chondroitin peaks in general, implying that used dermatan used in these experiments is somewhat more heterogeneous and contains more low affinity sites compared to the chondroitin.

Additional small peaks can occasionally be observed after the main peak for the chondroitin and dermatan columns. These peaks likely correspond to the binding of dimeric HGF-N that binds with higher affinity than the monomer as described in section 4.3. Analogous dimeric peaks can also be seen for the heparin columns, although these are much smaller. In this case, the same batch of protein was used, but for the chondroitin and dermatan affinity chromatography, the protein had been left at room temperature for a longer time than for the heparin affinity chromatography.

Elution of HGF-NK-C from the different GAG affinity columns is very similar to the behaviour of HGF-N. As the main GAG binding site is in the N-domain, the result is not surprising. HGF-NK-G follows the pattern of the two other protein constructs, the affinity for chondroitin A and C are similar and the affinity for dermatan sulfate is the lowest. For both types of heparin columns, there is a small shoulder at the beginning of the elution peak. This might be due to a heterogeneous protein sample or possibly due to a more complex binding event. HGF-NK-G elutes earlier than HGF-NK-C from both the in-house prepared and the standard GE Healthcare heparin affinity columns. As explained earlier, this is likely due to the different N-termini that these proteins have. A similar trend is observed for the binding to CSA, CSC and DS.

All protein constructs bound to sulfated GAGs (Figure 4.46). To investigate the importance of the sulfates and to see if HGF-N binds to non-sulfated GAGs, a hyaluronan affinity column was prepared in the same way as for the other GAG affinity columns. As can be seen in Figure 4.47, HGF-N interacts with hyaluronan, but elutes from the column approximately at physiological salt concentrations.

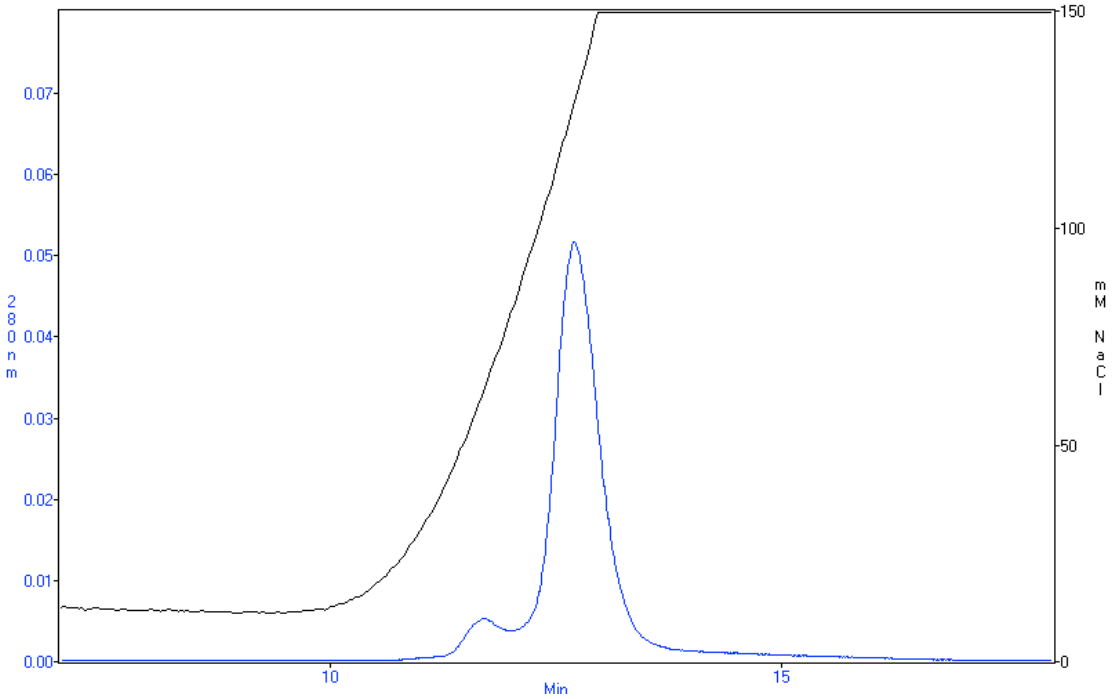


Figure 4.47 HGF-N eluted from a hyaluronan affinity column. The UV-trace is shown in blue and the conductivity is shown as mM NaCl in black. The protein binds to the column, but elutes at physiological NaCl concentration and is therefore unlikely to be physiologically relevant.

4.5.2 Affinity Chromatography of HGF-N with different N-Termini

To clarify the role of the N-terminal residues, the HGF-N proteins with different N-terminal kex2-cleavage sites described in section 4.2.1 were subjected to heparin affinity chromatography. The aim of the experiments was to understand how the N-terminal 6 residues contribute to the HGF-GAG interaction. The results are important in that they shed more light on the importance of the N-terminal residues that cannot be studied by NMR and x-ray crystallography.

Based on the fact that the N-terminus contains several basic residues, it was hypothesised that the more basic residues that are lacking from the N-terminus, the lower the heparin affinity and the earlier the proteins will elute from the heparin affinity column. This would also correspond to the order that the different proteins

elute from the MonoS cation exchange column during protein purification, as the heparin affinity is likely related to the pI of the protein.

It has previously been shown that the N-terminus of HGF is important for heparin binding [63]. The results further emphasise the importance of the N-terminal residues in the HGF-GAG interaction and although the residues are unstructured, they do indeed contribute to binding as shown by heparin chromatography (Figure 4.48).

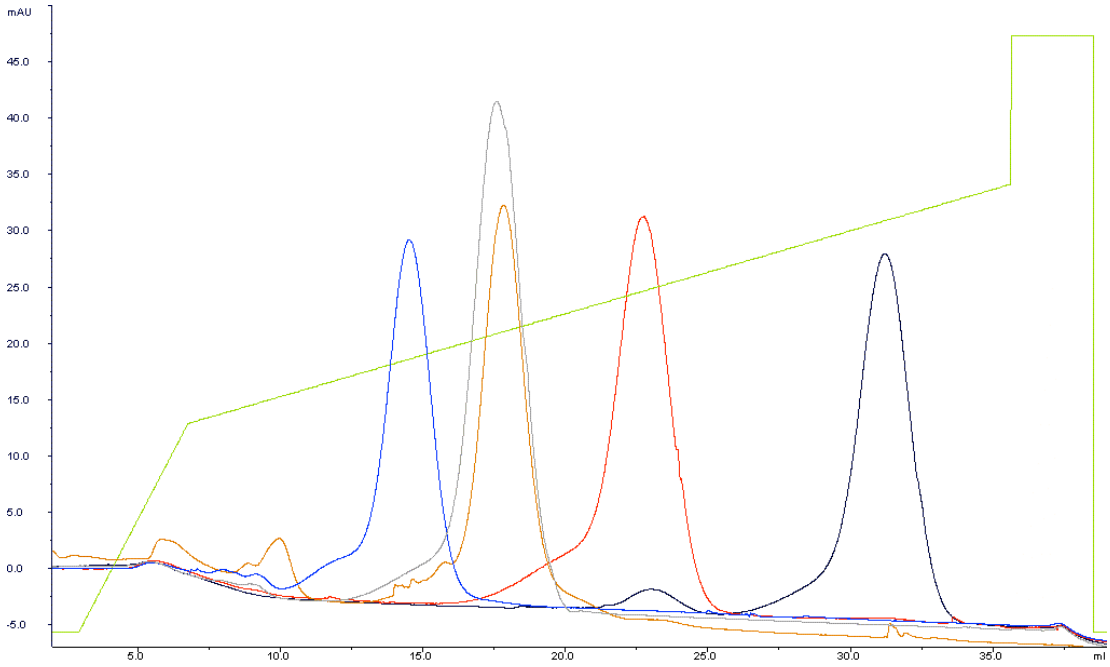


Figure 4.48 Heparin affinity chromatography of HGF-N with different N-termini as characterised in section 4.2.1. Blue trace: HGF-N-Δ6, lacking the first six residues (GQRKRR). Grey: HGF-N-Δ5, lacking the first five residues (GQRKR). Orange: S45 or T38 phosphorylated. Red: HGF-N. Black: Dimeric HGF-N.

The phosphorylation of S45 or T38 is expected to reduce GAG binding as it introduces a negatively charged, relatively big and bulky PO_4 group in the binding site. As only a small portion of the protein expressed in *P. pastoris* is phosphorylated, it cannot be correlated to human *in vivo* conditions, but it shows that

protein phosphorylation can be a way to regulate GAG binding *in vivo*. It is well established that protein phosphorylation is an important posttranslational modification for cell signalling in both extracellular and intracellular proteins. For example, it has been shown that neuroglycan c, a brain-specific transmembrane chondroitin sulfate proteoglycan can be phosphorylated [267].

Various protein constructs have been used for the structure determination of both free HGF and HGF-GAG complexes [21], [14], [17], [15]. The N-terminal residues have varied in these structures due to cloning procedures as summarised in Table 4.1.

Although it has been established that the N-terminal residues are important for the HGF-GAG interaction, these residues are disordered and it is therefore difficult to determine their structural role in the interaction. In the 2HGF NMR structure [17], the N-terminal residues are included, but inspection of the structure ensemble clearly shows that that the N-terminus is unstructured. In the 1GMO X-Ray structure [14], no electron density can be seen for the N-terminal residues and they are not included in the PDB file. The 1BHT X-ray structure determined by Ultsch *et al* [15] includes more N-terminal residues than the 1GMO PDB, but a 10 residue long N-terminal Flag epitope was included in this protein construct to facilitate purification, which likely disrupts the N-terminus.

In the HGF-NK-G – heparin structure (1GMO), the protein was crystallised with a 14mer heparin oligosaccharide, but only a maximum of 9 sugar units can be seen in the pdb file whereof four form the core binding site (Figure 4.49) [14], [2], [70]. The lack of sugar units at both ends of the heparin oligosaccharide is explained by poor electron density, as is evident from low B-factors in the pdb file [14]. Glycosaminoglycans are in general more flexible than proteins, but will to a certain extent be stabilised by protein binding. For the HGF-heparin interaction, it is likely that the GAG does binds across the N-terminus, but it cannot be visualised in the crystal structure due to low electron density.

It is also possible that steric hindrance and charge repulsion from the added N-terminal residues in the HGF-NK-G/1GMO protein construct (Table 4.1) distorts the native N-terminal protein-GAG interaction. Comparing different structures of the N domain, it looks like that the N-terminal residues reach towards the GAG and likely interact with it, becoming more structured upon binding (Figure 4.49). This theory is supported by the heparin chromatography results, the occurrence of new peaks in ^{1}H-^{15}N HSQC spectra when adding GAGs as discussed in section 4.10 and by the general importance of the N-terminal residues [63].

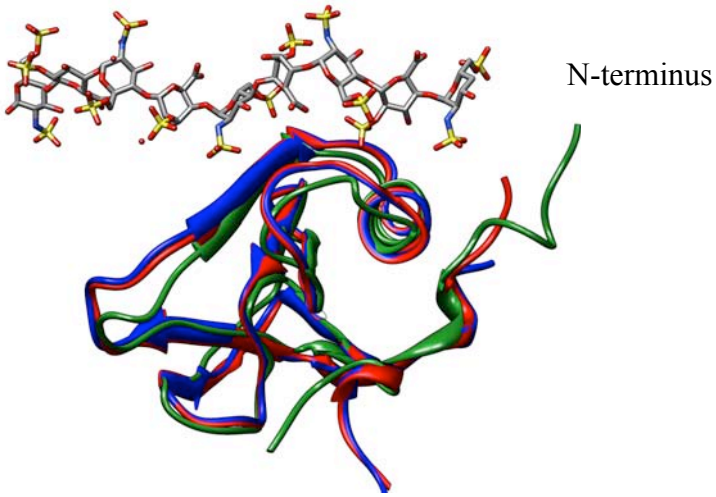


Figure 4.49 The PDB files 1BHT (Red), 1GMO (Blue) and 2HGF (Green) were overlaid and rendered in UCSF Chimera. Only the 1GMO PDB file contains heparin.

4.5.3 HGF Affinity Chromatography

Different GAG binding proteins often display a preference for a certain GAG sequence, although many different sequences often can be accommodated. It has previously been suggested that HGF binds preferentially to highly sulfated GAGs, just like many other GAG-binding proteins [69]. To investigate which heparin hexasaccharide sequence that binds tightest to HGF-N in a mixture of enzymatically digested heparin hexasaccharide, an HGF-N affinity column was produced by biotinylating the protein and immobilising it on a streptavidin column. The results are important because they objectively determine the affinity order of a heparin

hexasaccharide mixture and that HGF-N binds with the highest affinity to fully sulfated heparin hexasaccharide containing iduronic acid. Further, it is the first time that the current straightforward experimental setup is used to study the HGF-GAG interaction.

A GE Healthcare HiTrap streptavidin column containing streptavidin immobilised on Sepharose beads was used for the experiment. A PEG_4 linker was in place between the biotin group and the protein in order to minimise any sterical hindrance. A heparin DP6 mixture was loaded onto the column and eluted with a salt gradient, monitoring A_{232}. The last eluting peak should be the tightest binding GAG. Figure 4.50 summarises the overall concept of the experiment.

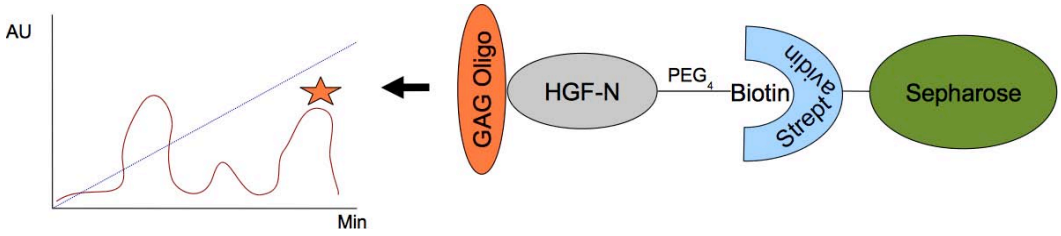


Figure 4.50 Outline of the HGF-N affinity experiment. Biotinylated HGF-N was immobilised on a streptavidin column. A heparin DP6 mixture was loaded onto the column and eluted with a NaCl gradient. The last eluting peak was judged to be the oligosaccharide binding with highest affinity.

There are several different reactive groups that can be employed to biotinylate proteins [268]. In this experimental setup, either the lysine/arginine and N-terminal amines or the aspartic/glutamic acid and C-terminal carboxyls could be targeted. As electrostatic sulfate - lysine/arginine interactions are among the most important points for protein-GAG interactions [269], aspartic/glutamic acid carboxyls are the preferred biotinylation target in order not to affect GAG binding at first glance. However, after carefully studying the structure of the HGF-N-GAG complex, it becomes clear that targeting the carboxyls would attach biotins in the binding site.

Targeting the amines with NHS chemistry is very efficient and straightforward. Keeping this in mind, it was decided to target the amines for biotinylation rather than the carboxyls, but to protect the binding site with heparin oligosaccharides to allow fewer biotins to attach there (Figure 4.51).

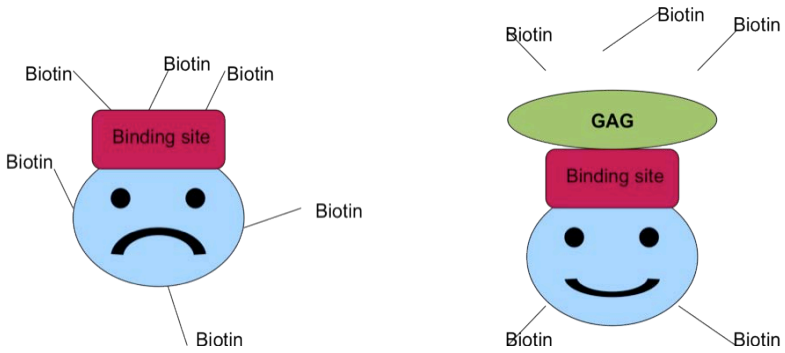


Figure 4.51 Protecting the binding site against biotinylation. HGF-N is symbolised by a blue smiley face with the binding site in red.

Although the binding site was protected by carrying out the biotinylation reaction in the presence of tenfold excess heparin DP8, it was expected that some biotin molecules would attach to the binding site and hinder the GAG interaction. To purify active HGF-N that bound heparin with normal affinity, the protein was heparin-affinity purified after the biotinylation reaction. Different ratios of the biotinylation reagents and protective heparin were explored (Figure 4.52). It was found that a twofold excess of biotinylation reagent and a tenfold excess of protective heparin gave good results.

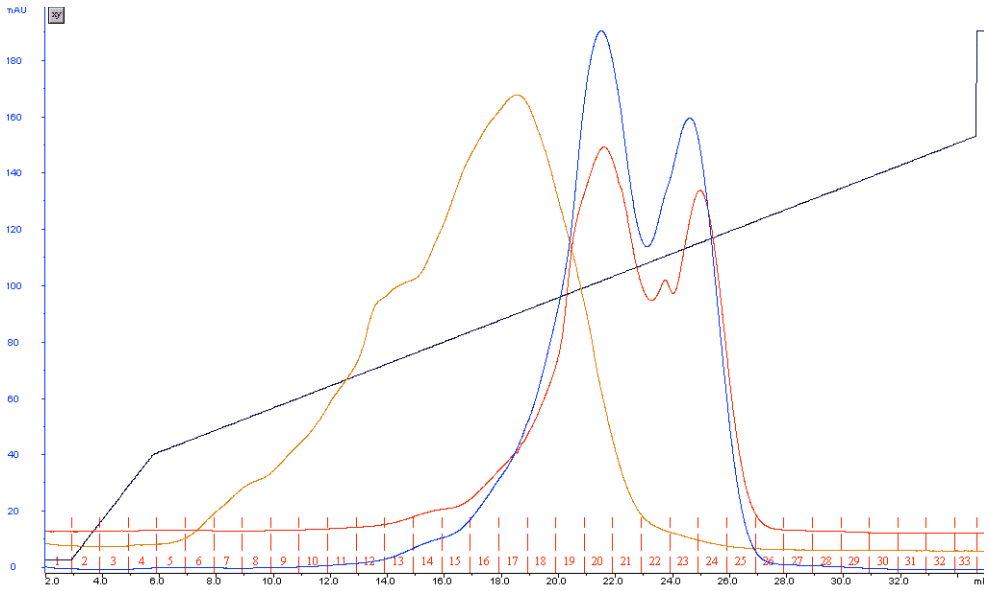


Figure 4.52 Heparin affinity purification of biotinylated HGF-N. The different chromatograms correspond to different reaction conditions as follows: Orange: tenfold excess of biotinylation reagent, fivefold excess of heparin DP8. Red: twofold excess of biotinylation reagent, fivefold excess of heparin DP8. Blue: twofold excess of biotinylation reagent, tenfold excess of heparin DP8.

The fractions of heparin affinity purified biotinylated HGF-N were analysed by MALDI-TOF mass spectrometry to assess how many biotin molecules were attached. The mass spectra show how the early eluting, low heparin affinity fractions in Figure 4.52 tend to have more biotin molecules attached than the later eluting high heparin affinity fractions. Fraction 17 in Figure 4.52 has 1, 2, 3, or 4 biotins attached, fraction 20 has 0, 1, 2 or 3 biotins attached, fraction 23 has 0, 1 or 2 biotins attached and fraction 25 has 0 or 1 biotin attached (Figure 4.53). These results show that even one biotin is enough to impede heparin binding if the biotin molecule is attached to the heparin binding site.

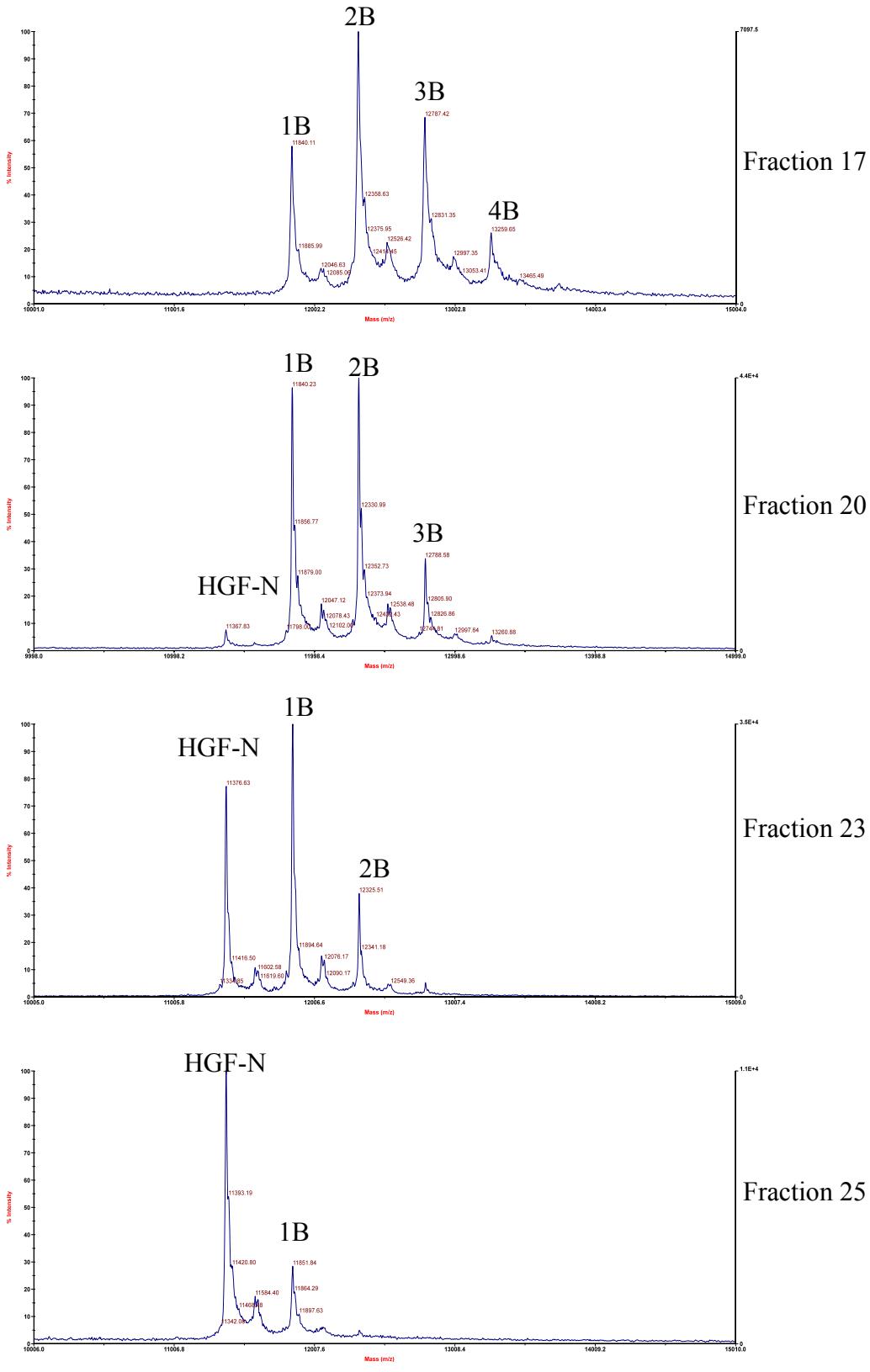


Figure 4.53 MALDI-TOF analysis of biotinylated HGF-N fractions from Figure 4.52. B: Biotin.

Based on the information obtained from the heparin affinity chromatography and the MALDI-TOF analysis, it was decided that fractions 21 to 26 in Figure 4.52 would be used for producing the final HGF-N affinity column.

To perform HGF-N affinity chromatography, approximately a twofold excess of size-exclusion purified heparin DP6 over HGF-N was loaded onto the HGF-N-affinity column and eluted with a salt gradient, monitoring A_{232} (Figure 4.54). As expected, the chromatography peaks are broad but distinguishable. Five different fractions were collected and analysed by SAX chromatography and MALDI-TOF mass spectroscopy.

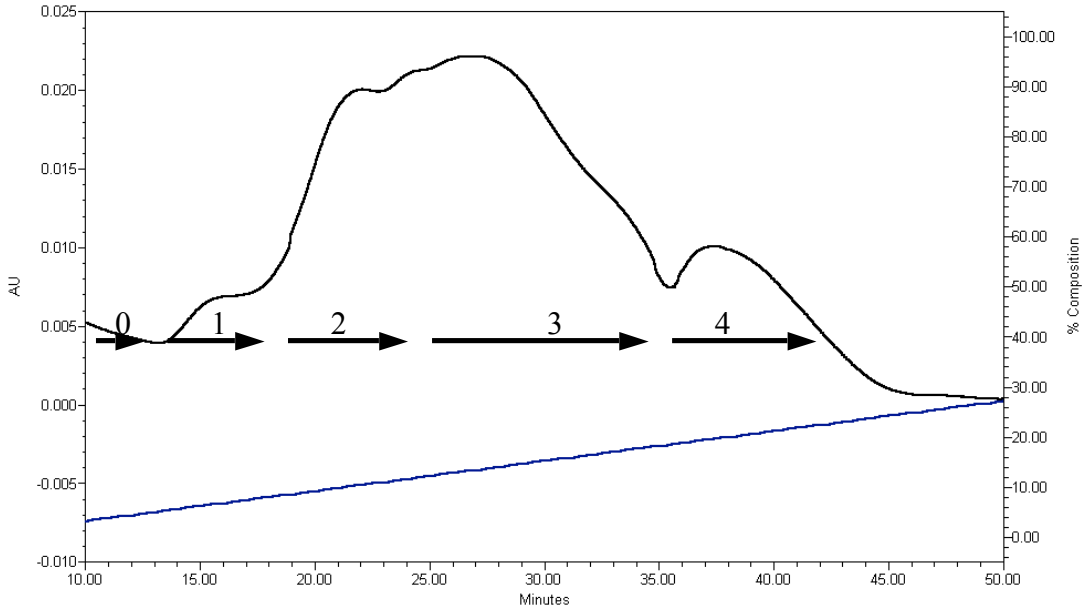


Figure 4.54 HGF-N affinity chromatography of heparin DP6. Fractions: 0) 0-13 minutes, 1) 14-18 minutes, 2) 19-25 minutes, 3) 26-34 minutes, 4) 35-45 minutes.

As expected, the affinity-purified heparin DP6 fractions elute late from the SAX column compared to the heparin DP6 mixture that the affinity-fractions were purified from (Figure 4.55). From experience, the SAX chromatograms indicate that the tightest binder (fraction 4) is fully sulfated heparin hexasaccharide containing iduronic acid residues and that the second tightest binder (fraction 3) lacks a sulfate or contains glucuronic acid motifs.

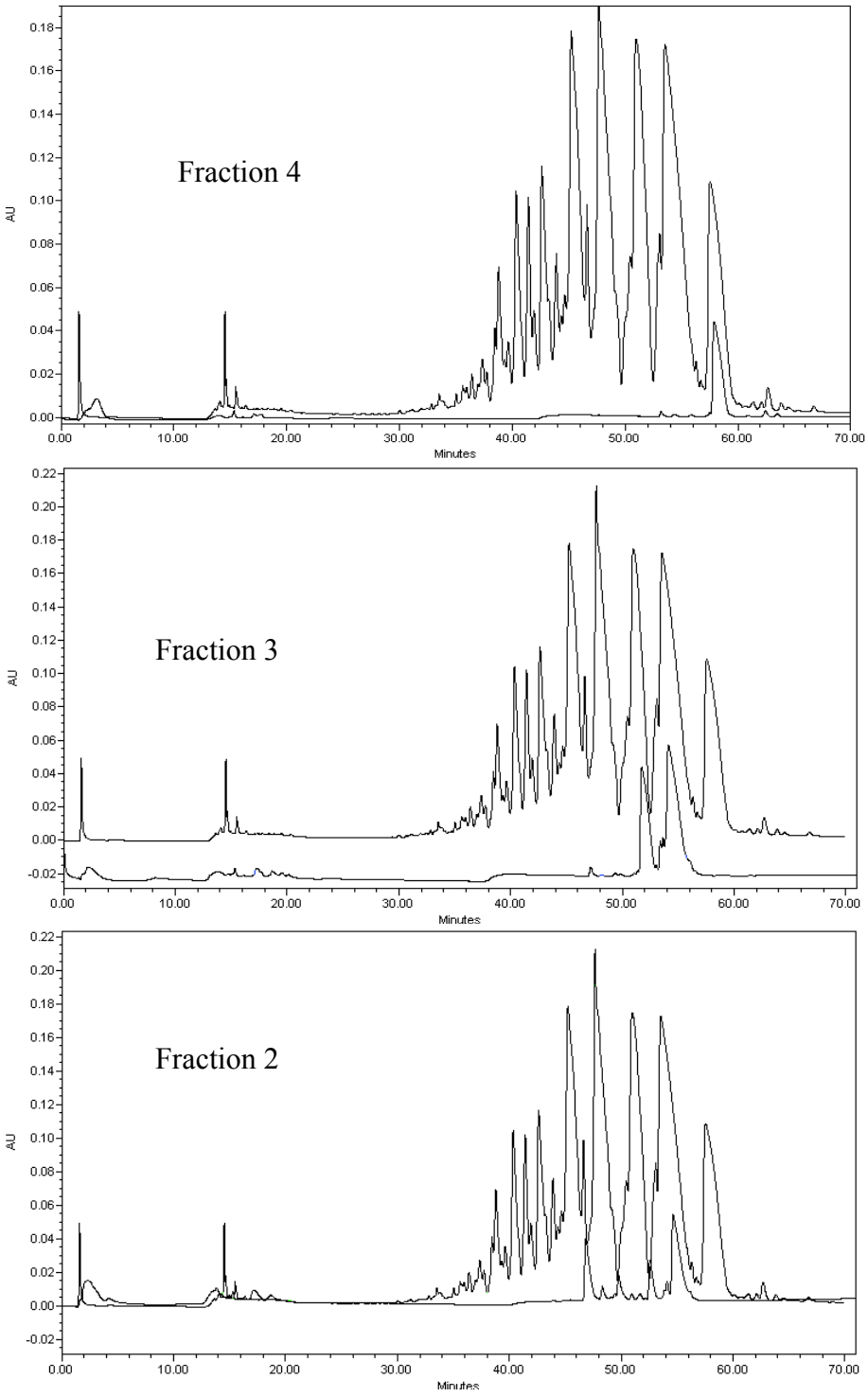


Figure 4.55 SAX chromatography of fractions from Figure 4.54. For comparison, the SAX chromatogram of the heparin DP6 mixture that was loaded onto the HGF-N affinity column is overlaid.

The SAX fractions were subsequently desalted and subjected to MALDI-TOF mass spectrometry to determine the degree of sulfation of the species. The MALDI-TOF spectra were calibrated with insulin and insulin b chain internal standards to improve accuracy. The standards contain several Na^+ adducts, although that does not impede spectrum quality (Figure 4.56). Both the free $(RG)_{11}$ and $(RG)_{11}$ + GAG complex can be identified, but most of the $(RG)_{11}$ peptide is complexed with the GAG. A small peak at 77 M/Z below the $(RG)_{11}$ + GAG complex indicates loss of one sulfate during the ionisation process. The calculated theoretical mass for fully sulfated protonated heparin DP6 is 1730.9 Da. The experimental mass from the spectrum in Figure 4.56 is 4095.95 - 2367.27 = 1728.65 Da.

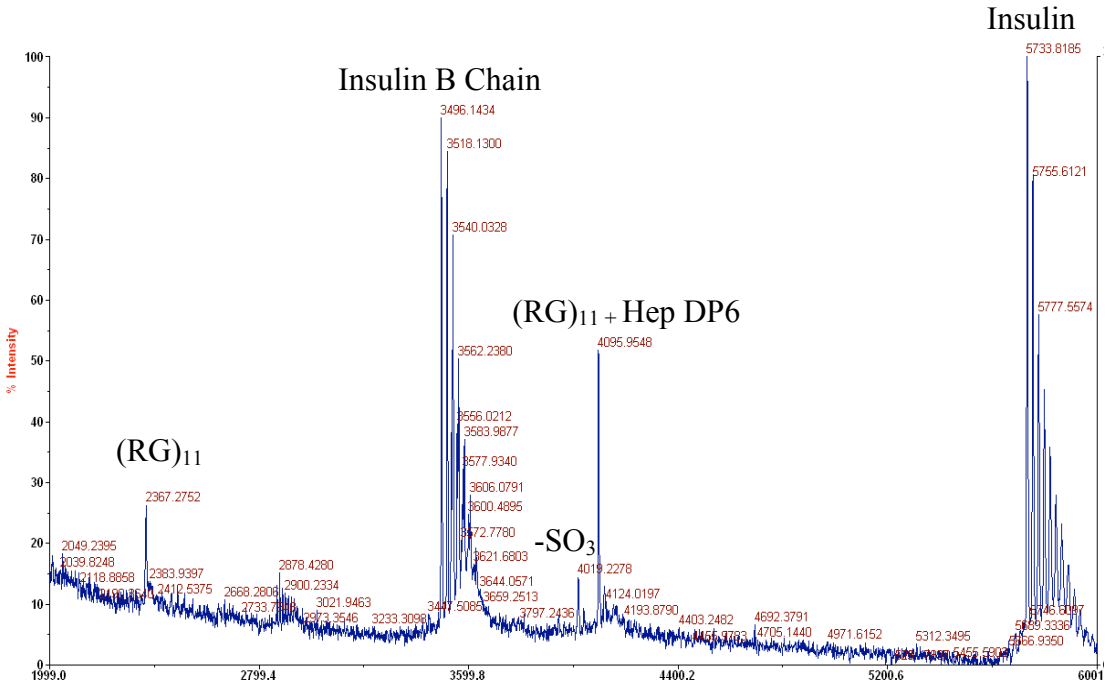


Figure 4.56 MALDI-TOF spectrum of peak 4, fully sulfated heparin DP6, complexed with $(RG)_{11}$ peptide. M/Z assignments: 2367.27: free $(RG)_{11}$ peptide, 3496.14: insulin b chain standard with Na^+ adducts, 4019.22: $(RG)_{11}$ + heparin DP6 that has lost one sulfate during ionisation, 4095.95: $(RG)_{11}$ + fully sulfated heparin DP6 and 5733.81: insulin standard with Na^+ adducts.

The spectrum in Figure 4.56 has not been deisotoped. With the employed experimental setup, reasonable isotope resolution can usually be achieved, but not for all samples (Figure 4.57).

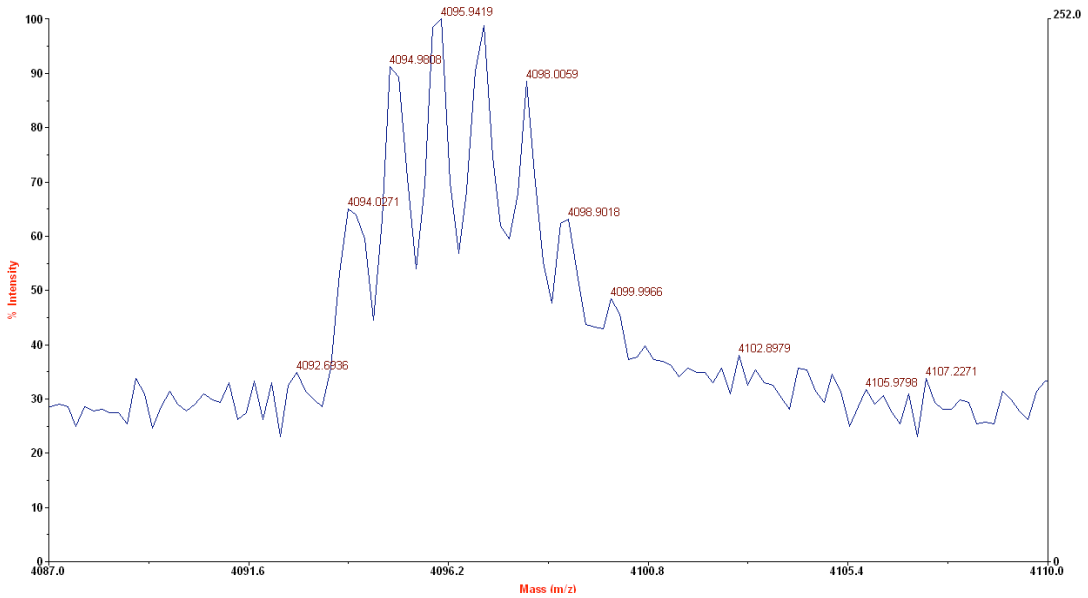


Figure 4.57 Zoom in of a $(RG)_{11}$ + heparin DP6 peak showing isotope resolution.

For spectra with poor isotope resolution, deisotoping is difficult. Even with relatively good quality spectra, deisotoping does not always work well. Figure 4.58 shows a deisotoped spectrum.

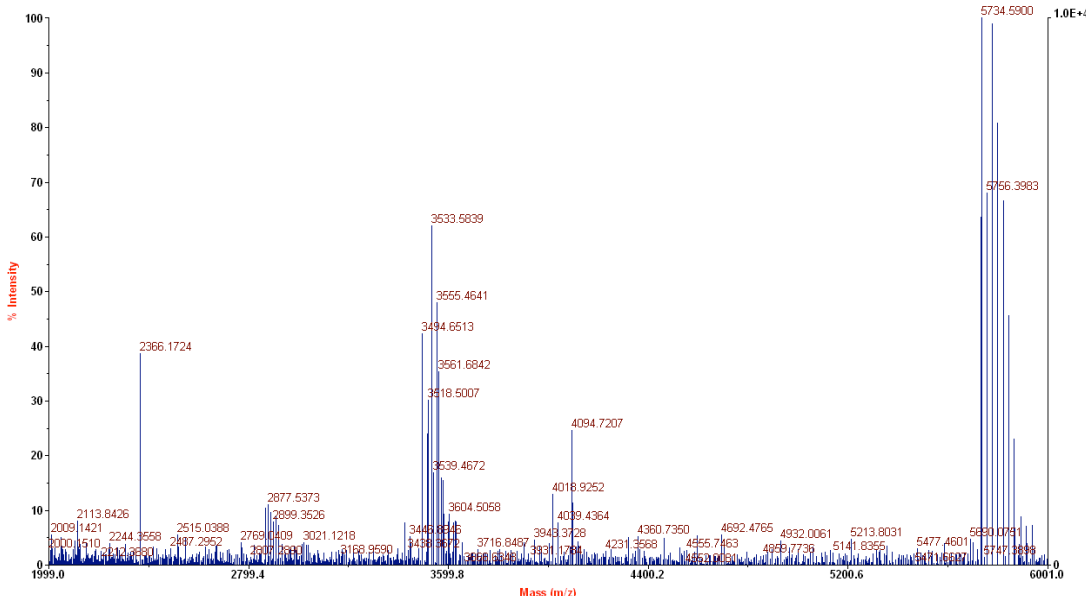


Figure 4.58 Deisotoped and calibrated $(RG)_{11}$ + heparin DP6 spectrum

The results from the SAX and MALDI-TOF analysis are summarised in Table 4.4.

Table 4.4 Summary of MALDI-TOF and SAX analysis of the different HGF-N affinity purified heparin DP6 oligosaccharides in Figure 4.54.

Fraction	**Molecular Weight**	**SAX**
0	Lack 2 or more sulfates	Many species
1	Lack 1 and 2 sulfates	3 Species
2	Lack 1 and 2 sulfates	4 Species
3	Lack 1 sulfate	2 Species
4	Fully sulfated DP6	1 Species

It is likely that two of the species in fraction 2 are the same as in fraction 3 as they have the same MW and elute from the SAX column at the same retention time (Figure 4.55).

As expected, the tightest binding heparin hexasaccharide is fully sulfated, consisting of IdoA2S - GlcNS6S disaccharides. The HGF-GAG interaction is mainly charge driven [14], as most other protein-GAG interactions [269]. Affinity experiments with Fondaparinux containing an extra 3-O sulfate described in section 4.8 showed that it bound very tightly and in general the more sulfates, the tighter protein-GAG interaction [269]. It was hypothesised that the tightest binding heparin DP6 oligosaccharide could contain a 3-O sulfation. The reason why a 3-O sulfation was not found could be due to:

1) The amount of 3-O sulfated heparin was too low to be detected on the HGF-N affinity column.
2) The used batch of heparin did not contain any 3-O sulfations or only in very low quantities.
3) A 3-O sulfation is not in the tightest binding GAG motif.
4) The 3-O sulfate motif was in the fully sulfated peak 4 fraction, but after SAX chromatography, it was not detected by MALDI-TOF due to very low absorbance (Figure 4.55).

A limitation of the current experimental setup might be the relatively low sensitivity of the employed UV-detection. The extinction coefficient of enzymatically digested GAGs at 232 nm is 5200 M^{-1} cm^{-1}, which is relatively low compared to many other fluorescent detection methods [168]. Attaching a fluorescent reporter would enhance sensitivity, but at the same time modify the GAG and thus potentially alter the protein-GAG interaction. An alternative approach could be to use ^{3}H labelled oligosaccharides and detect radioactivity, a technique that has been employed in similar experiments [66], [12].

4.6 Gel filtration Studies of HGF-N and HGF-NK-G Complexed With Different GAGs

It has been suggested that HGF dimerises/oligomerises upon addition of heparin/HS and possibly other GAGs that are long enough [2]. Most of the evidence comes from X-ray crystallography [14], crosslinking [265], [90] and gel filtration [14], [42], [16], [67] studies. In addition, it is a general mode of action of receptor tyrosine kinase ligands [18]. For example, the FGF-GAG complex is a well-characterised archetype of how a heparin binding growth factor oligomerises upon addition of the correct GAG motif [28], [270]. Although the stoichiometry of the HGF-GAG complex has been investigated before, it has not been as well studied as many other protein-GAG complexes, such as the FGF-GAG complex and the gel filtration studies of HGF-GAG complexes have not taken the big hydrodynamic radius of the GAG into as careful consideration as some studies of the FGF-GAG complex [28], [29], [270], [271].

It is likely that the HGF-GAG complex can oligomerise in several different ways and four possible different models (A-D) are shown in Figure 4.59. The most simple oligomerisation mode (A) is like beads on a string where one HGF molecule has one GAG binding site and if the GAG is long enough, many HGF molecules can bind. One can also imagine that HGF has one GAG binding site (B) and one HGF-HGF interface, this model is supported by the fact that HGF is reported as a dimer/oligomer in crystal structures and when it binds to GAGs [14]. Further, HGF could dimerise with the GAG between the two HGF molecules (C), the GAG being the spread that makes a double sandwich stick together. The most complex mode of dimerisation would be a combination of HGF-HGF interactions and a GAG sandwich model (D). Taking the Met receptor into consideration, these models are likely to be even more complex.

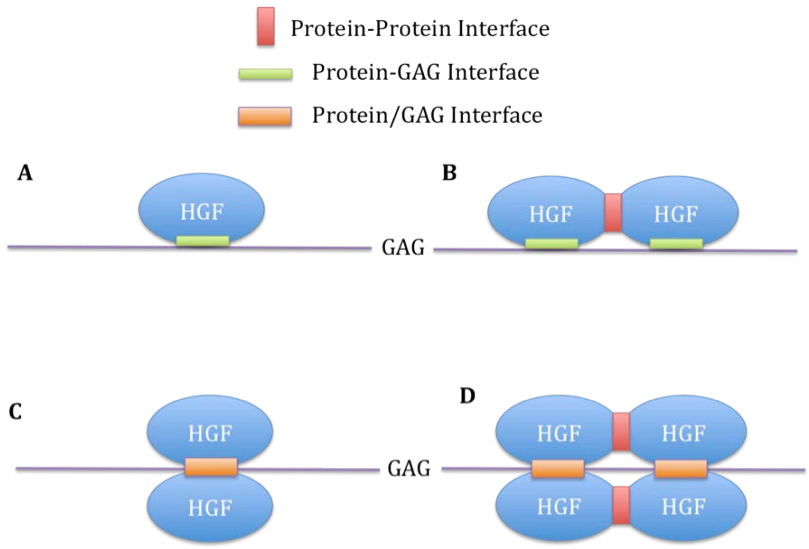


Figure 4.59 Four modes of oligomerisation of HGF-GAG complexes.

To further study the oligomerisation and binding process, different oligosaccharide lengths of CS, DS and heparin GAGs were mixed with HGF-NK-G and HGF-N at different ratios and studied by gel filtration. The aim of the experiments was to 1) determine the shortest GAG fragment that makes the proteins oligomerise, 2) clarify the oligomerisation mode, 3) take the large hydrodynamic radius of GAGs into careful consideration. The results are likely to be important in that it is the first time that a range of oligosaccharide lengths of three different GAG types, in particular CS, has been used to study the oligomeric mode of HGF-N and HGF-NK-G. The shortest fragment that causes HGF-N and HGF-NK-G to oligomerise and the oligomerisation mode of the proteins is currently not known.

Gel filtration is a common, simple and straightforward experiment to investigate the oligomeric state of proteins in solution that has been employed to clarify a wide range of biochemical problems. By using a gel filtration column that has been calibrated with protein standards in the investigated MW range it is possible to deduce the apparent molecular weight of the investigated species. In the current experimental setup, standards corresponding to monomeric and dimeric protein were used. The experiment design means that if the protein dimerises upon addition of

GAG, the apparent molecular weight as determined by gel filtration would double if the MW of the GAG can be ignored. The direct correlation between MW and elution volume on a gel filtration column depends on the hydrodynamic radius of the species and not the actual MW and thus the method is only reliable for globular proteins unless the column has been calibrated for other species.

To clarify the individual contributions of the GAG and the protein to the retention time shifts, the different GAGs and proteins were injected separately. To be able to monitor the UV-trace of the GAGs, the BioCAD FPLC system was used and the UV detector was set to 232 nm.

Results of the separate injections are summarised in the upper chromatogram in Figure 4.60. In these experiments CSC was used as a representative GAG. CSC DP12 migrates through the gel filtration column as a species with an apparent larger molecular weight than HGF-N and CSC DP22 migrates as a species with an apparent larger molecular weight than HGF-NK-G.

The middle chromatogram in Figure 4.60 shows how heparin DP12 migrates as a smaller apparent MW than HGF-NK-G, but when bound to the protein, the complex migrates as a broad peak with a large apparent MW, the excess of free DP12 GAG is also be visible. The broadness of the peak can be explained by a monomer / oligomer equilibrium, or due to the protein sample being heterogeneous, containing species with different GAG affinity.

The bottom chromatogram Figure 4.60 shows how heparin DP8 migrates at a larger apparent MW than HGF-N, the complex migrates as a narrow symmetrical peak and the free GAG can be seen. The MW of heparin DP8 is 2680 Da, the MW of HGF-N is 11370 Da, emphasising how quickly the elongated GAGs migrate through a gel filtration column compared to globular proteins. Although HGF-N is four times larger than heparin DP8, the latter still elutes earlier from the gel filtration column.

The tailing of the free protein peaks cannot be explained as no tailing is seen when running identical experiments on the Akta FPLC system.

These experiments have shown that an increase of the apparent molecular weight of a protein-GAG complex is strongly affected by the large hydrodynamic radii of GAGs. It is difficult to determine the oligomeric state of a protein-GAG complex unless modelling of the hydrodynamic radii of the complexes is undertaken. Modelling of the complex requires knowledge of the geometry and strength of the complex. However, if the results are interpreted with care, keeping the hydrodynamic radius of the GAG in mind, some conclusions can be drawn.

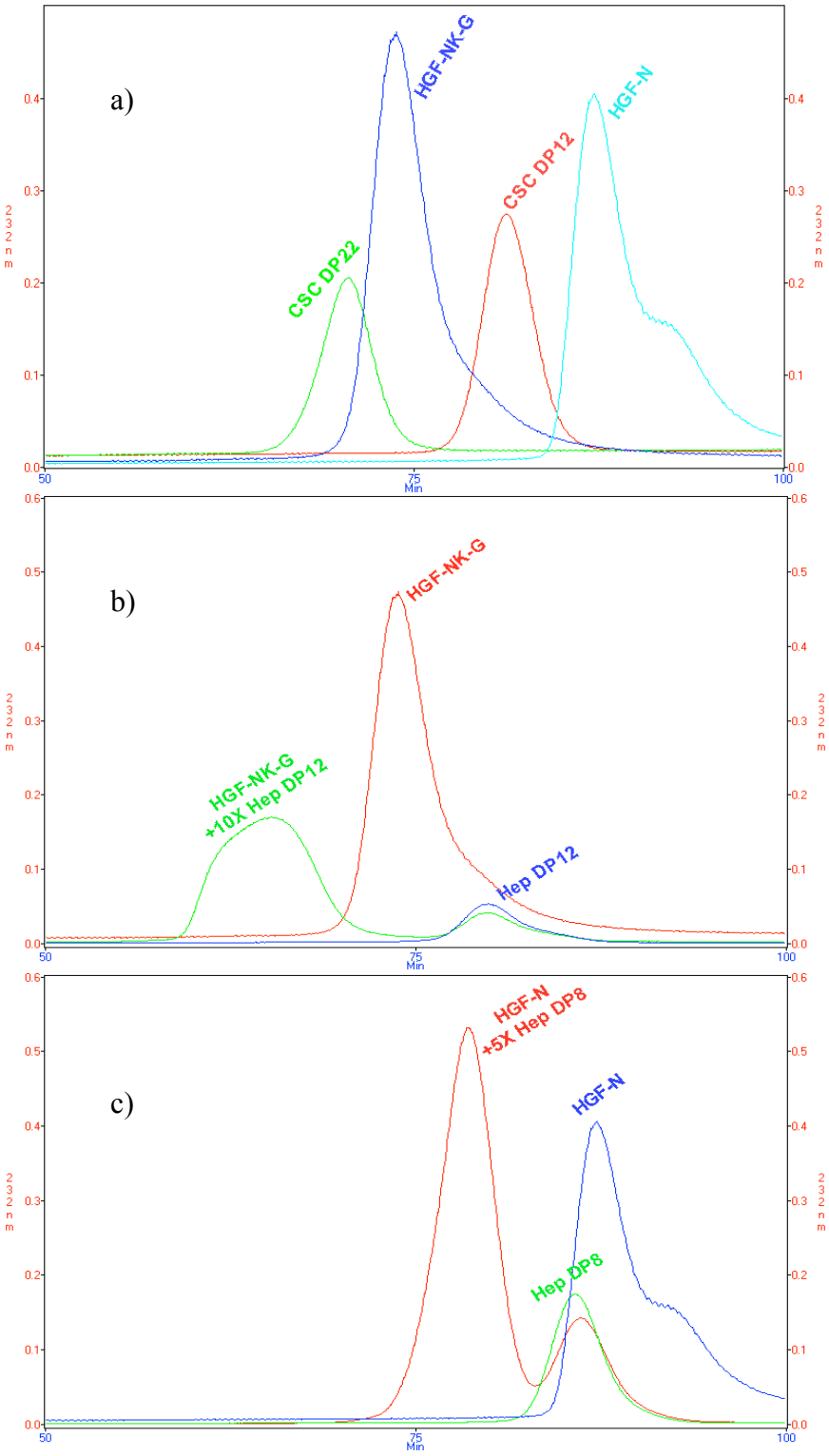


Figure 4.60 Gel filtration of different HGF constructs and GAGs in PBS. a) All four species monitored in separate runs. b) Free HGF-NK-G, free heparin DP12, HGF-NK-G + heparin DP12 in tenfold excess mixture. c) Free HGF-N, free heparin DP8 and HGF-N + heparin DP8 in tenfold excess mixture.

To investigate the oligomerisation and binding differences between HGF-N, HGF-NK-G and different GAG oligosaccharides, the proteins were mixed with GAGs of varied oligomer lengths at a 1:1 ratio (unless otherwise stated) and analysed by gel filtration monitored at 280 nm, which detects the protein but not the GAG. Similar experiments have been reported with Human Complement Factor H and dextran sulfate [272] and many other protein-GAG complexes, including HGF.

Initially, HGF-N was mixed with DP4-DP16 heparin oligosaccharides and analysed by gel filtration (Figure 4.61). It can be seen that all oligosaccharides bind to HGF-N and induces elution volume shifts. The octasaccharide peak is broadened, the DP10 and DP14 peaks have shoulders that disappear with DP16 at tenfold excess.

There are several possible explanations for the appearance of shoulders for DP10 and DP14, the two most appealing are:

1) A DP10 GAG is long enough to accommodate two protein molecules and equilibrium is reached where some DP10 molecules have two protein molecules bound (1:2) and some have one protein molecule bound (1:1).
2) The factions are not clean and contaminated with longer oligosaccharides.

These results support the beads on a string oligomerisation model in Figure 4.59 for HGF-N as no sharp shift in elution volume is seen and the elution volume shifts are likely due to the addition of more saccharide units. At tenfold excess DP16, some higher order complexes can be seen in the chromatogram: the UV absorbance increases before giving rise to the main symmetrical peak.

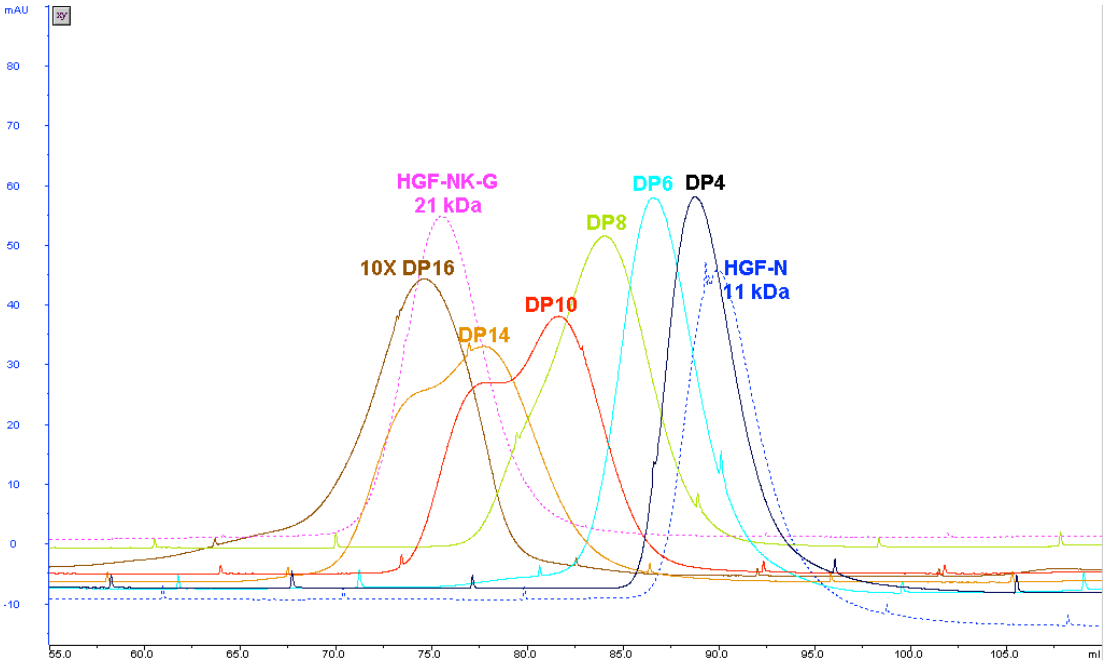


Figure 4.61 Gel filtration of free HGF-N, free HGF-NK-G as a standard and HGF-N in the presence of heparin oligosaccharides. The UV-absorbance was monitored at 280 nm.

CSC with a lower sulfate and IdoA content is expected to bind HGF-N with much lower affinities than heparin. This is indeed the case as CSC mixture elutes mostly as free protein and the UV-traces are elongated (Figure 4.62). This suggests that at an equimolar ratio, the binding site is far from saturated. However, at tenfold excess of CSC DP20, only one symmetrical peak is visible.

As discussed in section 4.5.1, there are few, if any reports at all on that HGF interacts with CS under physiological conditions. These gel filtration studies of HGF-CS complexes complements the results in section 4.5.1 and further shows that it is likely that HGF interact with CS under physiological conditions, although with much lower affinity than for heparin and HS.

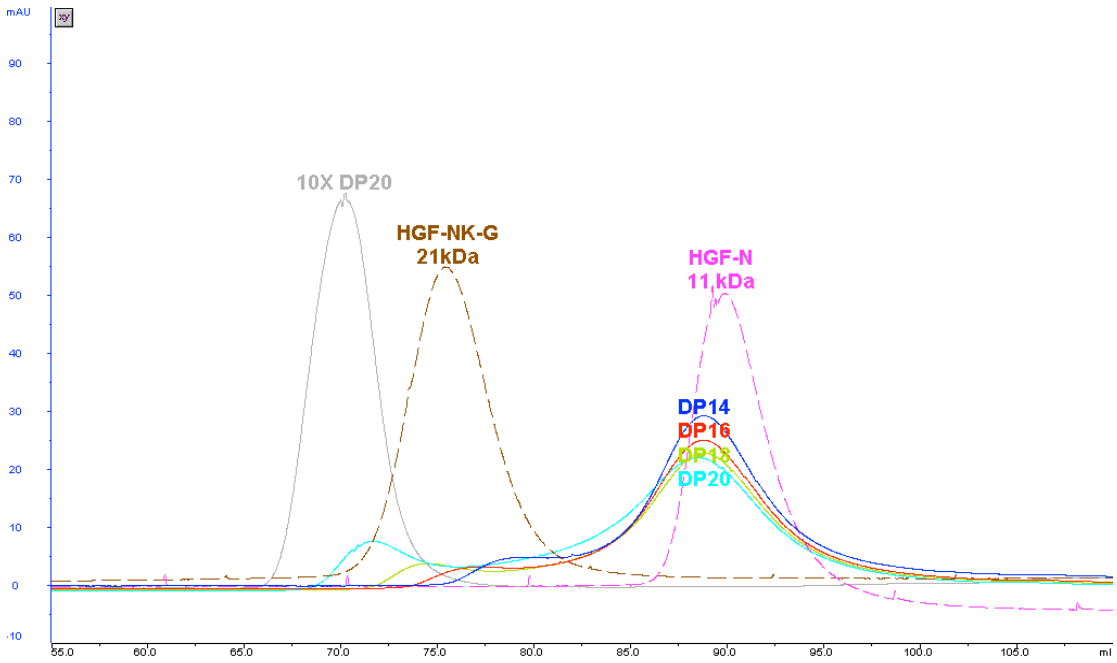


Figure 4.62 Gel filtration of free HGF-N, free HGF-NK-G as a standard and HGF-N in the presence of CSC oligosaccharides. The UV-absorbance was monitored at 280 nm.

The UV-traces for the HGF-N + DS mixtures are elongated and depending on the length of the DS oligosaccharide, a different proportion of the protein is in the free or bound form (Figure 4.63). For longer oligosaccharides, the elution UV profile consist of two peaks and for an increase in oligosaccharide length, a corresponding increase in apparent molecular weight is observed. One peak appears at the same elution volume as free protein and the other peak corresponds to the complex that gradually shifts towards higher apparent molecular weight with increasing oligosaccharide length. The splitting of peaks is interpreted as the lower apparent molecular weight corresponds to free protein and that the higher apparent molecular weight corresponds to the protein-GAG complex. This behaviour is similar to that of CSC.

DS with a lower sulfate content than heparin, but with a higher IdoA content than CSC is expected to bind to HGF-N with affinities between CSC and heparin. This is also the case, as the bound form is more prevalent in the chromatogram, but the

results do not correlate with Figure 4.46, where it was found that HGF-N and HGF-NK-G bind tighter to CS than to DS. This is probably due to that different sources of DS were used for the different experiments. The DS for this experiment was from Grampian Enzymes whereas the DS for the affinity chromatography was from Malcolm Lyon, University of Manchester.

Interestingly, the elution profiles for heparin versus CSC/DS are different. The heparin peaks do not show free protein and only the complex peak can be seen. This is likely due to that HGF-N binds heparin with higher affinity.

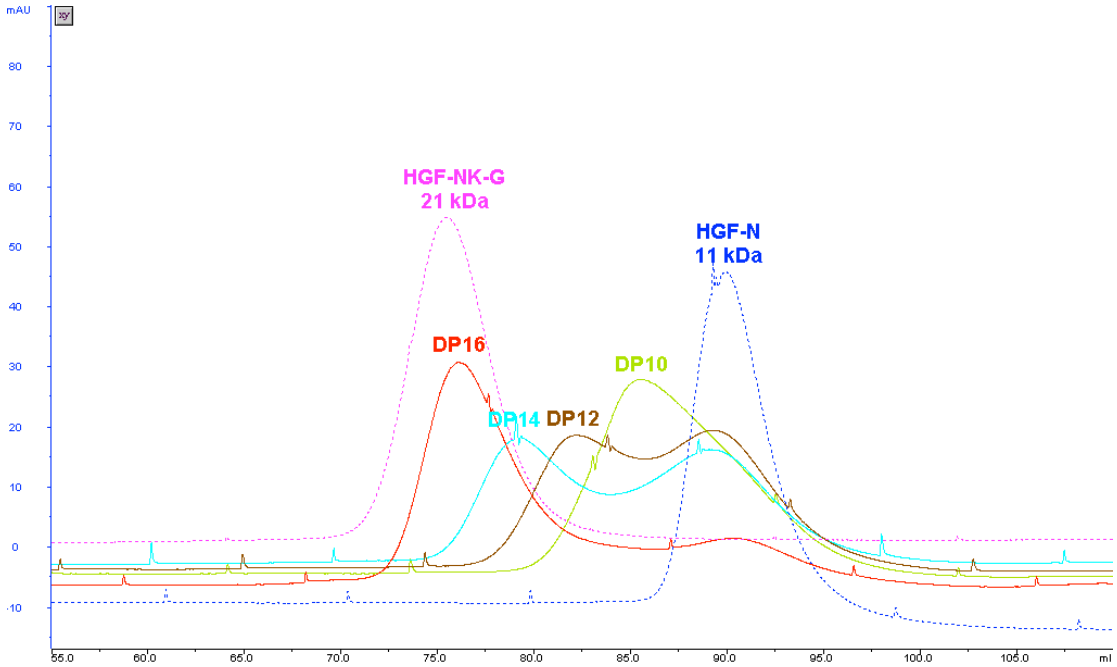


Figure 4.63 Gel filtration of free HGF-N, free HGF-NK-G as a standard and HGF-N in the presence of DS oligosaccharides. The UV-absorbance was monitored at 280 nm.

Comparison of HGF-N in the presence of DP10 and DP14 fractions of heparin, CSC and DS is particularly interesting. Heparin shows two peaks eluting at 73 and 78 minutes which likely corresponds to 2:1 and 1:1 protein:GAG complexes. CSC also

shows two peaks eluting at 78 and 88 minutes, which corresponds to the 1:1 complex and free protein. The same applies to DS. These observations strongly support the A and B models in Figure 4.59. Only heparin binds tightly enough and has a high enough exothermic contribution to compensate for the loss in entropy associated with binding two protein molecules.

To investigate the oligomerisation and binding properties of HGF-NK-G, the protein was mixed with different GAGs and analysed by gel filtration in a similar manner as for HGF-N.

The gel filtration chromatograms of HGF-NK-G mixed with different heparin oligosaccharides exhibit similar behaviour as HGF-N. For shorter heparin oligosaccharides, an elution time shift can clearly be seen due to the larger hydrodynamic radius of the complex (Figure 4.64). A clear double peak can be seen for heparin DP16, comparable to the DP10 and DP14 results for HGF-N. This can be explained by that HGF-NK-G is a larger protein and a longer oligosaccharide is needed to host two protein molecules.

Heparin DP16 was used in equimolar and tenfold excess ratios. The two chromatograms overlay well, but at tenfold excess, the 1:1 peak is larger than the 1:2 peak. This is expected as the larger excess of heparin DP16 favours a 1:1 complex.

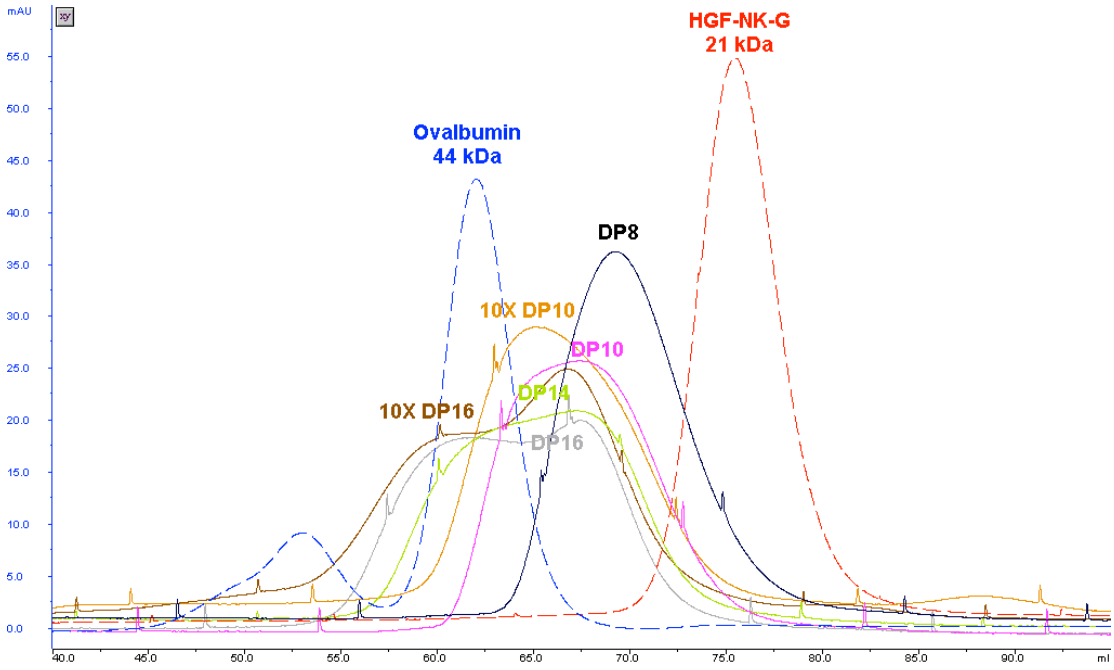


Figure 4.64 Gel filtration of free HGF-NK-G, free ovalbumin as a standard and HGF-NK-G in the presence of heparin oligosaccharides. The UV-absorbance was monitored at 280 nm.

HGF-NK-G binds very weakly to CSC as can be seen in Figure 4.65. Also, the apparent MW of the complexes seem smaller compared to those for the heparin oligosaccharides. This is likely due to that the protein spends more time uncomplexed than complexed. The effect can also be partly explained by that the heparin is more highly sulfated, which likely increases the hydrodynamic radius. HGF-NK-G complexed with CSC DP20 at tenfold excess migrates approximately at the same rate as HGF-NK-G complexed with equimolar heparin DP10.

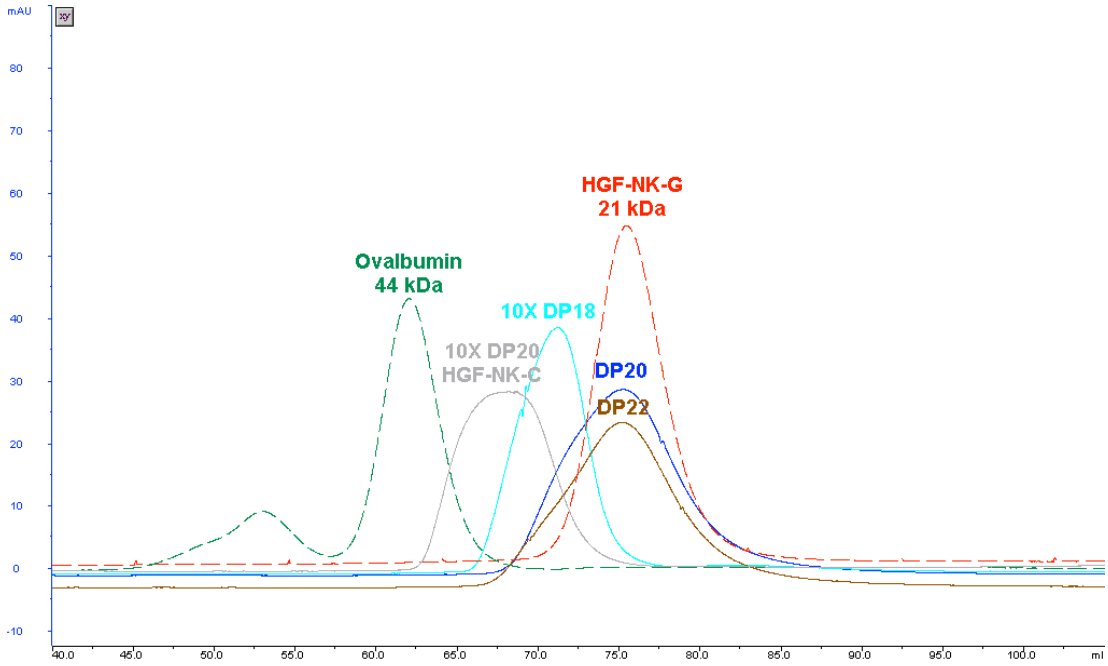


Figure 4.65 Gel filtration of free HGF-NK-G, free ovalbumin as a standard and HGF-NK-G in the presence of CSC oligosaccharides. HGF-NK-C was used for the 10X DP20 chromatogram. The UV-absorbance was monitored at 280 nm.

The results for HGF-NK-G complexed with DS are similar to those for CSC. Interestingly, the tenfold excess DS DP20 chromatogram in Figure 4.66 contains a third higher order shoulder that is not visible for CSC, this could possibly be two proteins on one GAG oligosaccharide or a DS impurity.

Just like for HGF-N, no 2:1 complexes could be seen for HGF-NK-G in the presence of CSC or DS oligosaccharides, only heparin oligosaccharides showed 2:1 complexes. As expected, HGF-NK-G due to its larger size needed a longer oligosaccharide than HGF-N to form a 2:1 complex. HGF-N formed 2:1 complexes with heparin DP10, whereas HGF-NK-G formed 2:1 complexes with heparin DP14 (significantly) and DP16 (clearly).

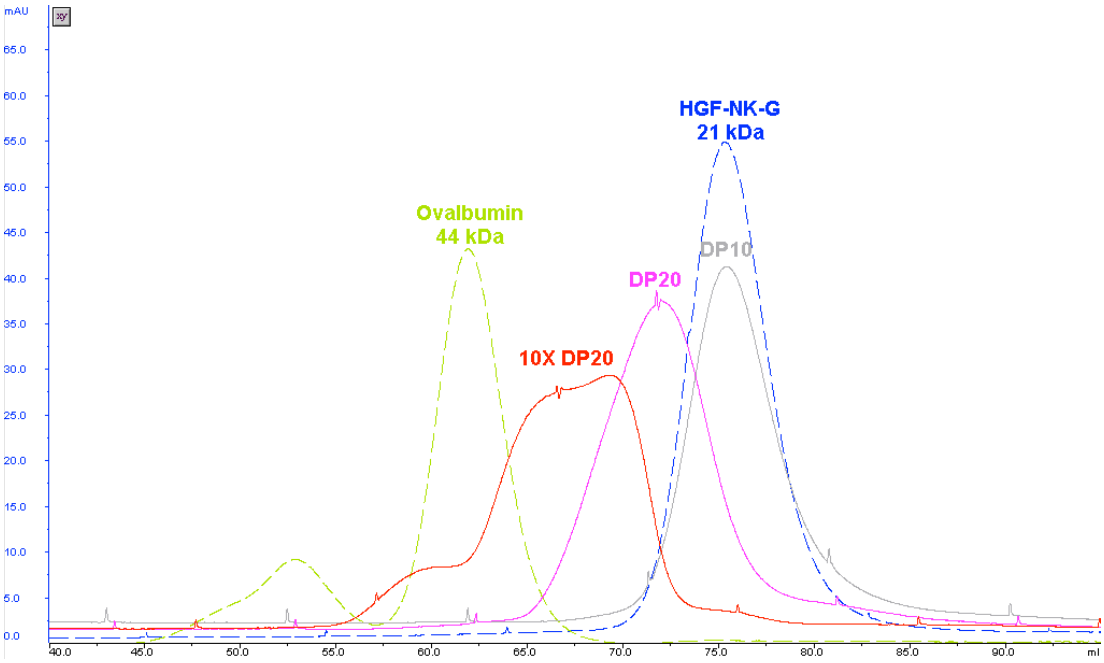


Figure 4.66 Gel filtration of free HGF-NK-G, free ovalbumin as a standard and HGF-NK-G in the presence of DS oligosaccharides. The UV-absorbance was monitored at 280 nm.

In conclusion the gel filtration experiments showed that:

1) When interpreting gel filtration data of protein-GAG complexes, it is necessary to take the large hydrodynamic radius of the GAG into account.

2) Evidence for binding of two protein molecules on one GAG oligosaccharide was obtained for heparin DP>10 for HGF-N and DP>14 for HGF-NK-G, but not for CSC or DS oligosaccharides. This suggests that HGF-N and HGF-NK-G oligomerises like beads on a string in the presence of heparin (mode B in Figure 4.59)

3) It is unlikely that that these complexes have a *trans* orientation (Figure 4.59) because:

    a. A *trans* orientation would also be possible for shorter GAGs, for which ternary complexes were not observed.

b. 2:1 complexes are only seen when the GAG oligosaccharide has sufficient length.

c. HGF-NK-G requires a longer GAG than HGF-N

4) It was not possible to determine if the interaction was cooperative and includes formation of protein-protein interfaces.

In the current experiments, the column was not calibrated and the hydrodynamic radii of the complexes were not calculated and the results were interpreted using elution times only. With the structures for HGF-NK-G complexed with heparin available, PDB files for different complexes can be created and the hydrodynamic radii calculated with programs such as HYDROPRO [233] and CRYSOL [273]. This treatment could potentially lead to a more sophisticated analysis of the data.

The gel filtration column used in the current experiments is a preparative column, leading to a big waste of material. A Superdex 75 10/300 GL or a Superdex 75 PC 3.2/30 column would be better suited as it would use significantly less material and shorten run times.

An alternative method to determine stoichiometry of protein-GAG complexes that overcomes the problem of the large hydrodynamic radius of carbohydrates is mass spectrometry and it has been successfully employed to characterise the FGF-heparin complex [28]. The most physiological way to carry out mass spectrometry for these complexes is to use an electrospray ionisation interface and a volatile buffer such as ammonium acetate or ammonium bicarbonate, but MALDI ionisation can also be used as shown in this project (Figure 4.86). However, mass spectrometry is clearly not a solution-based method and artefacts can easily arise and as often when studying protein-GAG interactions by mass spectrometry, the heterogeneity of the GAG oligosaccharides pose a problem. This is especially a problem in mass spectrometry experiments with relatively long heparin-derived oligosaccharides that can have a very high degree of heterogeneity, as the m/z spectrum will be crowded with broad

peaks stemming from different substitutions and adducts. Ideally, a GAG species purified to homogeneity should be used, but it is difficult to achieve with DP > 8 due to the resolution of SAX columns and loss of material. Although, the loss of material might not pose a big problem for mass spectrometry, as it is a very sensitive method. Sometimes alternative GAG purification methods can be employed to purify homogeneous species for mass spectrometry studies such as protein-GAG affinity chromatography, or mixing the GAG and protein and injecting the mixture on a gel filtration column and using the gel filtration fractions that correspond to the complex MW, as the low affinity GAGs will have diffused away from the protein during the gel filtration run. The heterogeneity also poses problems when using mixtures for other techniques such as NMR, SPR, ITC, affinity chromatography and AUC as some species in the mixture might not bind, some species bind with intermediate affinities and some species bind tightly.

Alternative techniques to study stoichiometry of protein-GAG complexes are ITC that gives the stoichiometry of the complex and SPR where the response can be correlated to how many protein molecules bind to an immobilised ligand. AUC can also be used to study the stoichiometry of biomolecular complexes and is generally a more advanced and powerful technique compared to gel filtration, but the derived information is still similar and the large hydrodynamic radius of carbohydrates can again make it difficult to use for protein-GAG interactions.

4.7 Analytical Ultracentrifugation

Gel filtration results in section 4.6 suggested that HGF-N and HGF-NK-G oligomerise as beads on a string. It is always a good idea to use an alternative technique to confirm experimental results. Therefore, AUC was carried out for HGF-N and HGF-NK-G complexed with Fondaparinux and heparin DP16 as a method to study oligomerisation in addition to gel filtration. The aim of the experiments was to determine the oligomeric state of the protein construct in their free and complexed states. The results are important in that they are the first of its kind for HGF-N and

HGF-NK-G and the support gel filtration results. AUC experiments were performed in collaboration with Alexandra Solovoya, University of Newcastle.

A typical moving boundary profile obtained in the experiments is shown in Figure 4.67.

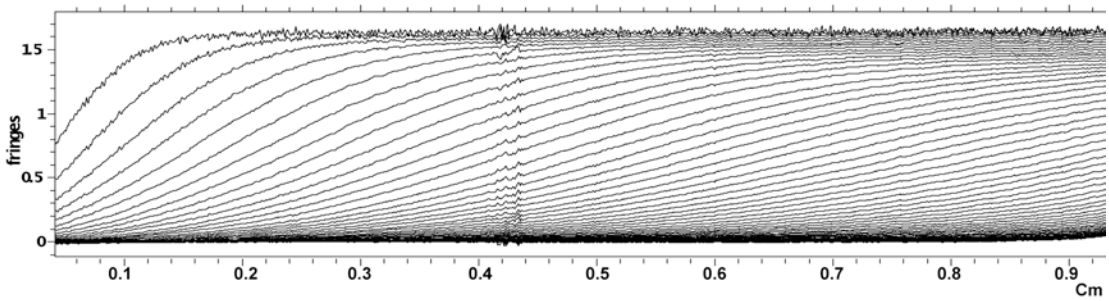


Figure 4.67 Example of a moving boundary profile of HGF-N complexed with Fondaparinux at an equimolar ratio using interference optics.

Distributions of sedimentation coefficients (c(s)) plotted vs sedimentation coefficient (Svedberg), show that both HGF-N and HGF-NK-G are monomeric in solution in the used concentration range (10-120 μM for HGF-N and 10-60 μM for HGF-NK-G) (Figure 4.68). When the proteins are complexed with Fondaparinux a sedimentation coefficient shift can be seen due to the larger hydrodynamic radius of the complex. A slight broadening of the peak can be noticed, probably due to the binding dynamics of the complex. The peak remains relatively sharp and symmetrical for HGF-N, indicative of a homogeneous ligand that does not cause oligomerisation while it is broader and asymmetric for HGF-NK-G. When the proteins are complexed with heparin DP16, a larger sedimentation coefficient shift can be seen and the peak becomes wider and more asymmetric, indicative of heterogeneity of the heparin DP16, oligomerisation and a dynamic equilibrium of the complex.

The linear extrapolation of the sedimentation coefficients to zero concentration for free HGF-N and HGF-NK-G in solution does not indicate formation of higher order species in solution as can be seen in Figure 4.69. There is no noticeable concentration dependence for the HGF-N + Fondaparinux complex in the examined protein concentration range. On the other hand, concentration dependence was observed for the HGF-NK-G + Fondaparinux complex, indicating a presence of some higher molecular weight species. As expected, concentration dependence of the sedimentation coefficient can be seen for both proteins complexed with heparin DP16, which is indicative of higher molecular weight species.

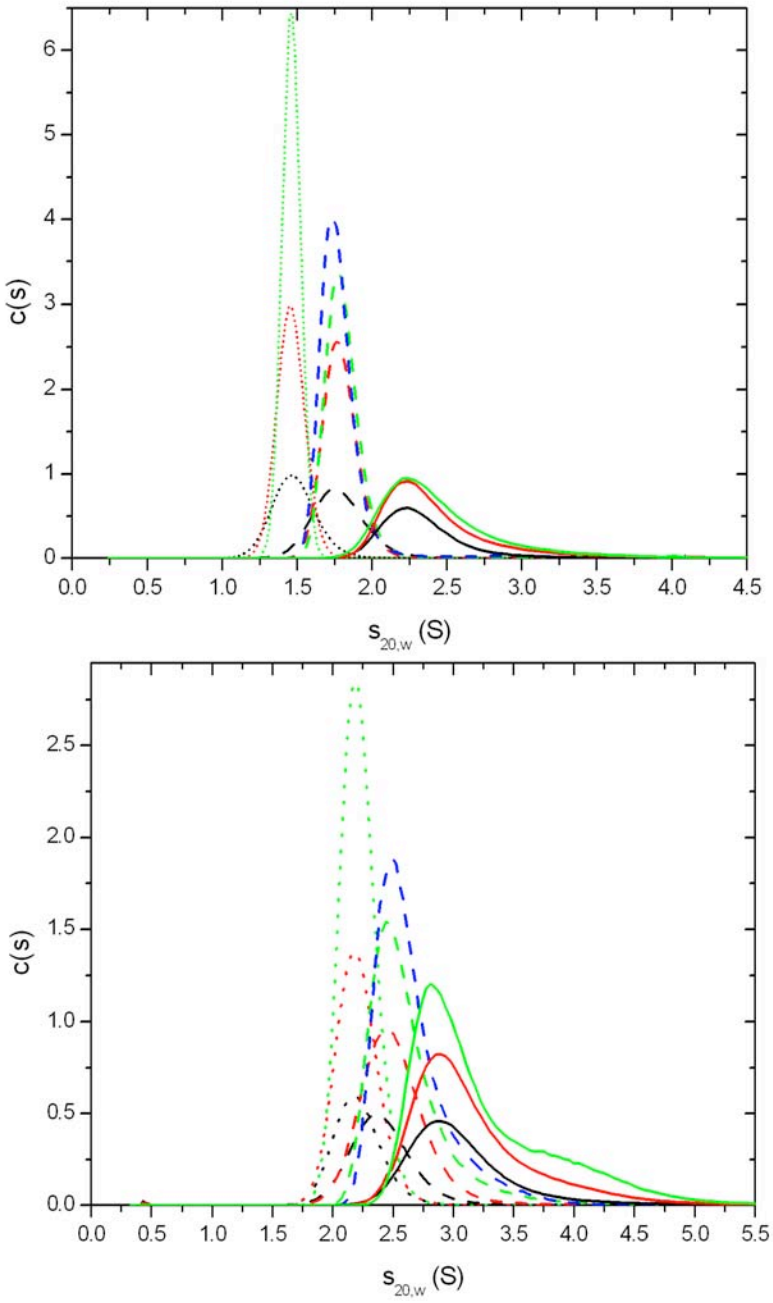


Figure 4.68 Distributions of sedimentation coefficients (c(s)) plotted vs sedimentation coefficient (Svedberg) for HGF-N (a) and HGF-NK-G (b) at different protein concentrations using UV absorbance detection. Dotted lines represent protein alone, dashed lines represent protein complexed with Fondaparinux at a 1:1 radio and solid lines represent protein complexed with heparin DP16 at a 1:1 ratio. The black trace represents the lowest concentration, red the medium concentration, green the higher concentration and blue the highest concentration. Sample concentrations can be found in Figure 4.69.

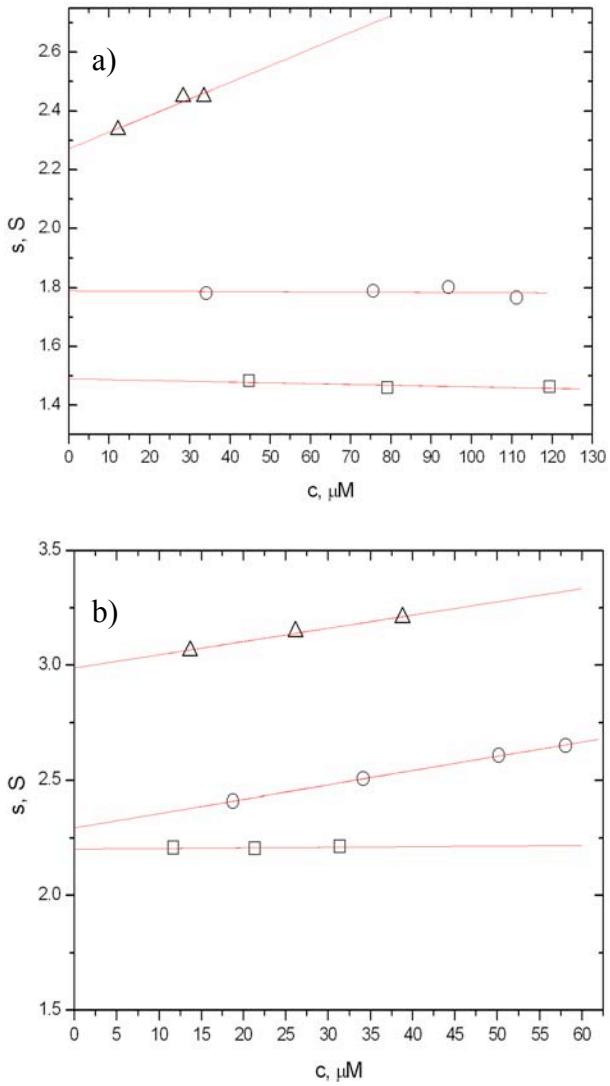


Figure 4.69 Linear approximation of sedimentation coefficients to zero concentration using absorbance data. a) HGF-N, b) HGF-NK-G. Symbols: □: free protein, ○: protein + Fondaparinux at 1:1 ratio, △: protein + heparin DP16 at 1:1 ratio.

The sedimentation coefficients and molecular masses for free HGF-N and HGF-NK-G correspond well with the calculated values for monomeric protein as can be seen in Table 4.5.

Table 4.5 AUC parameters derived from absorbance data.

Sample	MW (kDa)	s PDB file (S)	One-dimensional c(s) model		Two-dimensional c(s, f/f_0) model		
					Species 1	Species 2	Species 3
			$s_{20,w}$ (S)	M (kDa)	$s_{20,w}$ (S)	$S_{20,w}$ (S)	$s_{20,w}$ (S)
HGF-N	11.4	1.41	1.47±0.01	11.8±0.23			
HGF-N+Fx			1.78±0.02	13.4±0.24			
HGF-N +DP16					2.37±0.02	3.41±0.14	4.47±0.05
HGF-NK-G	21.1	2.17	2.21±0.01	21.7±0.82			
HGF-NK-G + Fx					2.52±0.04	3.44±0.09	
HGF-NK-G + DP16					3.02±0.04	4.10±0.1	5.47±0.23

HGF-N remains a homogeneous monomer when complexed with Fondaparinux, but the data is more complex for HGF-N + heparin DP16 and for HGF-NK-G + heparin DP16 and Fondaparinux. As can be seen in Figure 4.68 and Figure 4.69, these complexes show concentration dependence and their molecular weights cannot be determined. By employing an alternative data fitting method, it was possible to resolve several species for both the HGF-N and HGF-NK-G complexes. In this model, the frictional ratio (f/f_o) was used, which corresponds to the ratio of the experimental frictional coefficient of a molecule relative to its theoretical spherical frictional coefficient [274]. The frictional ratio reports on particle shape, but it has been shown that the shape dependence is relatively weak and spherical proteins typically have a frictional ratio of 1.2-1.3 whereas asymmetrical proteins typically have a frictional ratio of 1.5-1.8 [275].

The calculated c(s, f/f_o) distribution is shown as a two-dimensional distribution with grid lines representing the sedimentation coefficient and frictional ratio of the analysis. In the c(s, f/f_o) two-dimensional model of HGF-N complexed with heparin DP16, two to three species can be seen (Figure 4.70). These peaks could arise when several HGF-N molecules bind along the GAG chain, producing an elongated species as indicated by a high frictional ratio and a relatively small difference in

sedimentation coefficient. At an equimolar protein:GAG ratio, most of the protein binds in a 1:1 mode, while some 2:1 protein:GAG complex are also present. There are also two very small peaks indicating even higher order species. These could be artefacts from the data analysis or a due to 3:1 complex. Approximately 13.5% of the species seems to be present as a 2:1 protein:GAG complex.

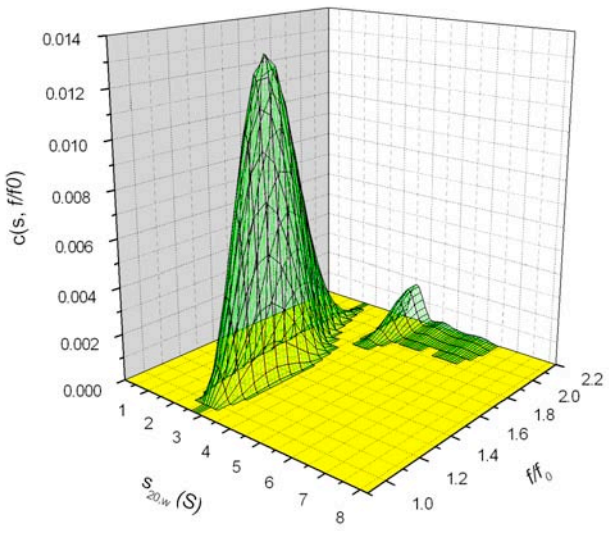
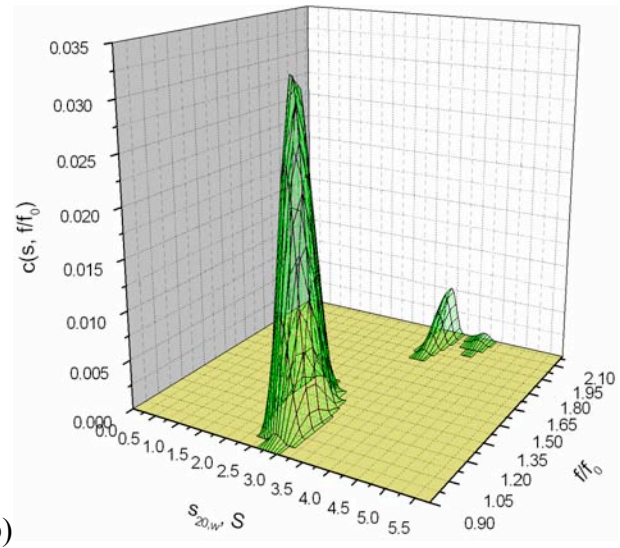


Figure 4.70 Two-dimensional distribution model c(s, f/f_o) analysis of HGF-N + heparin DP16 complexes.

Two peaks (one major and one minor) with higher sedimentation coefficients are also present in the two-dimensional c(s, f/f_o) analysis the HGF-NK-G + Fondaparinux complex (Figure 4.71), indicating that a small fraction of HGF-NK-G binds two Fondaparinux molecules per protein molecule, although the sample was analysed at a 1:1 ratio. Reassuringly, this has also been noticed in MALDI-TOF spectra with HGF-NK-G + Fondaparinux (data not shown). The additional Fondaparinux molecule could bind to the weaker heparin binding site in the K domain [14]. It is therefore possible that the AUC data also shows formation of such a complex. The heparin DP16 results for HGF-NK-G are similar to those for HGF-N with two or three higher order complexes present, probably indicating binding of several protein molecules to one GAG. Approximately 16.4% of the species seems to

be present in a 2:1 protein:GAG form. Interestingly, a shoulder can be seen on the main peak of HGF-NK-G complexed with heparin DP16. It is possible that this peak represents a different binding mode, forming a more globular complex than the other peaks that stem from more elongated species.

a)



b)

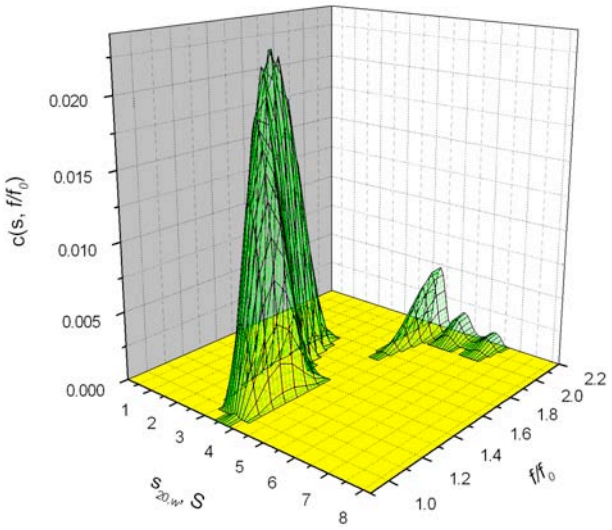


Figure 4.71 Two-dimensional distribution model c(s, f/f_0) analysis of HGF-NK-G complexes. a) HGF-NK-G + Fondaparinux, b) HGF-NK-G + heparin DP16.

The results obtained from the AUC experiments correspond well with the results from the gel filtration experiments. Higher-order complexes consisting of two proteins on one GAG chain could be seen for both HGF-N and HGF-NK-G in gel filtration experiments and AUC experiments. The special case of HGF-NK-G binding two Fondaparinux molecules was confirmed in MALDI-TOF experiments, but not by gel filtration as an equivalent gel filtration sample was not analysed.

4.8 Isothermal Calorimetry

Different experimental techniques for the determination of K_D values often yield different results. NMR experiments carried out in this project and preliminary ITC data obtained by Bärbel Blaum, University of Edinburgh, suggest that the HGF-GAG interaction is relatively weak with a K_D in the low micromolar range. FITC labelled heparin and fluorescence spectroscopy showed that the K_D of the HGF-N - heparin interaction is 1-2 μM [21]. On the other hand, SPR/biosensor analysis of the HGF-GAG interaction presented in several studies have reported K_D values in a low nanomolar range.

One of the first SPR studies of the HGF-GAG interaction used crude heparin that was biotinylated on the odd free amines of heparin in order to immobilise it on a BIAcore carboxymethyldextran - streptavidin surface. The study determined that the K_D for the full-length HGF – heparin interaction is 4-30 nM and the K_D for NK2 – heparin interaction is 2.3 nM [11]. A different study biotinylated the protein core of endocan, a CS/DS proteoglycan and immobilised it on an aminosilane-streptavidin surface determine the HGF - endocan interaction had a K_D of 1.8 - 9.2 nM [40]. Using different types of HS, Rahmoune *et al* found three different types of HGF binding sites in HS with different kinetics and K_Ds in the range of 200 pM to 3 nM. Rahmoune *et al* studied the interactions by biotinylating the HS on free amines and using a carboxymethyldextran - streptavidin chip [276]. Using Decorin biotinylated on the core protein, it was established that full-length HGF binds to DS with

affinities in the range of 19.7 nM [12]. The K_D of full-length HGF to *Ascidia nigra* DS that is 2O and 6O –sulfated is in the range of 29-37 nM [67].

Delehedde *et al* biotinylated heparin oligosaccharides on the reducing end and established that full-length HGF binds to heparin-derived DP4-DP14 oligosaccharides with similar K_Ds in a range of 9 - 16 nM [10]. Interestingly, no trend was observed indicating that longer oligosaccharides would bind tighter. Interpreting the study by Delehedde *et al*, heparin DP4 with a reducing end ring opened by attachment of a "XX" linker by reductive amination is enough to bind full-length HGF: the K_D for heparin DP4 was 11 nM, DP6 9.1 nM and DP14 16 nM.

A problem when studying protein-GAG interactions with SPR/biosensors is that GAGs are long and heterogeneous, making it hard to produce valid mathematical models. The mathematical model is complicated by the fact that there are different binding motifs with different K_Ds and dynamics and likely a number of protein molecules will bind to the same GAG chain, depending on its length. The higher the injected protein concentration, the more protein molecules will bind to the lower affinity GAG motifs, making it hard to saturate the surface, which is important for establishing a correct K_D. By developing the right model for the right system and by using an extensive data analysis, the problem of multiple proteins on one GAG chain can be overcome for some well-characterised interactions [277].

Furthermore, the negatively charged dextran surface used in many biosensor systems such as the BIAcore CM5 chip is not suitable to study interactions of an electrostatic nature such as protein-GAG interactions (Tim Fagge, GE Healthcare, personal communication). Often, alternative surfaces with less dextran are available (such as BIAcore CM3 and CM4), or with a plain carboxylated surface (such as BIAcore C1), which are more suitable for studying biomolecular interactions of an electrostatic nature.

Some protein-GAG complexes have very low K_Ds in a low nM range, for example the ATIII-pentasaccharide complex, bFGF, annexin II and V, whereas other protein-GAG and protein-carbohydrate interactions are in the μM range [69]. As pointed out above, different types of GAGs and immobilisation techniques have been used to study protein-GAG interactions by SPR. It has been shown that the protein-heparin binding kinetics depends on the heparin immobilisation technique [278]. Also, the determined kinetics varies with lab, type of instrument, experiment design and data analysis method [279]. The need of immobilisation of the ligand can affect the binding mode of the ligand and mass transfer can also affect the results.

ITC, on the other hand, has been described as the golden standard for characterisation of the strength of biomolecular interactions [280]. ITC has been used to study several protein-GAG interactions, for example the FGF-GAG interaction [30]. In a single ITC experiment, the thermodynamics of the binding reaction are characterised by determining the stoichiometry of the interaction (n), the association constant (K_A), the free energy (ΔG), enthalpy (ΔH) and entropy (ΔS). The main advantage of ITC over SPR is that it is a solution-based technique with no need for immobilisation and that it can obtain all thermodynamic parameters in one single experiment. A disadvantage of using ITC to study weak interactions is that a lot of material is required. This is because the heat change of the binding must be large enough to be measurable by the ITC experiment. The heat change depends on the enthalpy and strength of the interaction.

The aim of these experiments was to use ITC as a robust solution state technique to study the affinity, thermodynamics and stoichiometry of the HGF-GAG interaction. To my knowledge, there are no published studies of this kind in the literature. The results are important, as very few solution-based alternative techniques to SPR have been used to study the HGF-GAG interaction.

In order to get ITC c values that are high enough and yield reliable data for interactions with a K_D ~ 10 μM, approximately 100 μM protein has to be used. This corresponds to approximately 2-5 mg of HGF-N and HGF-NK-G. The quality of ITC data, especially the stoichiometry of the interaction, is highly dependent upon accurate determination of the protein and ligand concentration. In this study, the concentrations were carefully determined by UV-absorbance.

To compare the binding characteristics of different lengths of oligosaccharides to different protein constructs, HGF-N and HGF-NK-G was titrated with Fondaparinux, heparin DP6, DP10 and DP16. Control titrations were performed to establish the heat of dilution. Titration of CS and DS oligosaccharides into HGF-N was attempted, but sufficient responses were not achieved. Figure 4.72 shows examples of obtained ITC data and the results are summarised in Table 4.6.

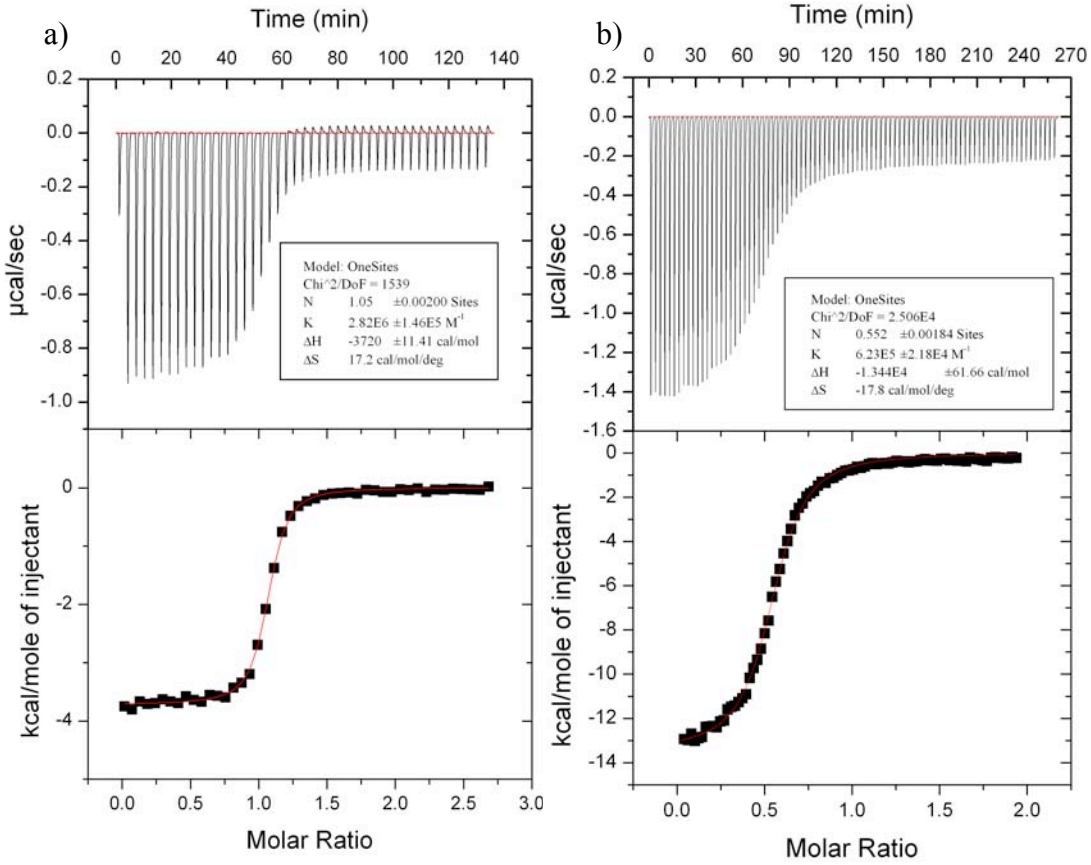


Figure 4.72 ITC data for HGF-N titrated with Fondaparinux (a) and with heparin DP16 (b). Molar ratio corresponds to GAG:protein ratio.

It should be emphasised that the oligosaccharides used in these experiments were only size fractionated and therefore represent mixtures of many species with different sulfation patterns. The ITC data can therefore be analysed only in a qualitative manner, as different species have different thermodynamic binding properties.

Table 4.6 ITC data for HGF-N and HGF-NK-G binding to different heparin oligosaccharides. Fx: Fondaparinux.

HGF-N				
DP	**Stoichiometry (Ratio)**	**Dissociation Constant (M)**	**Enthalpy (kJ/mol)**	**Entropy (J/mol/deg)**
16	0.56±0.003	1.76E-06±0.08E-06	-57.20±0.36	-78.24
10	0.81±0.002	2.30E-06±0.08E-06	-48.70±0.20	-52.72
6	0.82±0.004	1.41E-06±0.1E-06	-36.55±0.27	-8.58
Fx	1.05±0.002	3.55E-07±0.02E-06	-15.56±0.05	71.96
HGF-NK-G				
DP	**Stoichiometry (Ratio)**	**Dissociation Constant (M)**	**Enthalpy (kJ/mol)**	**Entropy (J/mol/deg)**
16	0.53±0.003	2.79E-06±0.13E-06	-55.48±0.44	-76.57
10	0.95±0.004	5.43E-06±0.19E-06	-43.68±0.26	-43.51
6	1.13±0.011	9.26E-06±0.60E-06	-28.62±0.38	1.98
Fx	1.15±0.006	5.71E-06±0.26E-06	-18.02±0.12	40.88

From the Gibbs free energy equation ($\Delta G = -RT \ln K = \Delta H - T\Delta S$), it is clear that a negative ΔH (enthalpy) and a positive ΔS (entropy) are indicative of a tighter interaction. At the same time, when a decrease in enthalpy is compensated by an increase in entropy a similar K_D is maintained. The data suggest that this indeed is

the case here and that the entropy contribution of the interaction becomes more favourable with decreased oligosaccharide length, while the enthalpy contribution becomes more favourable with increased oligosaccharide length (Figure 4.73).

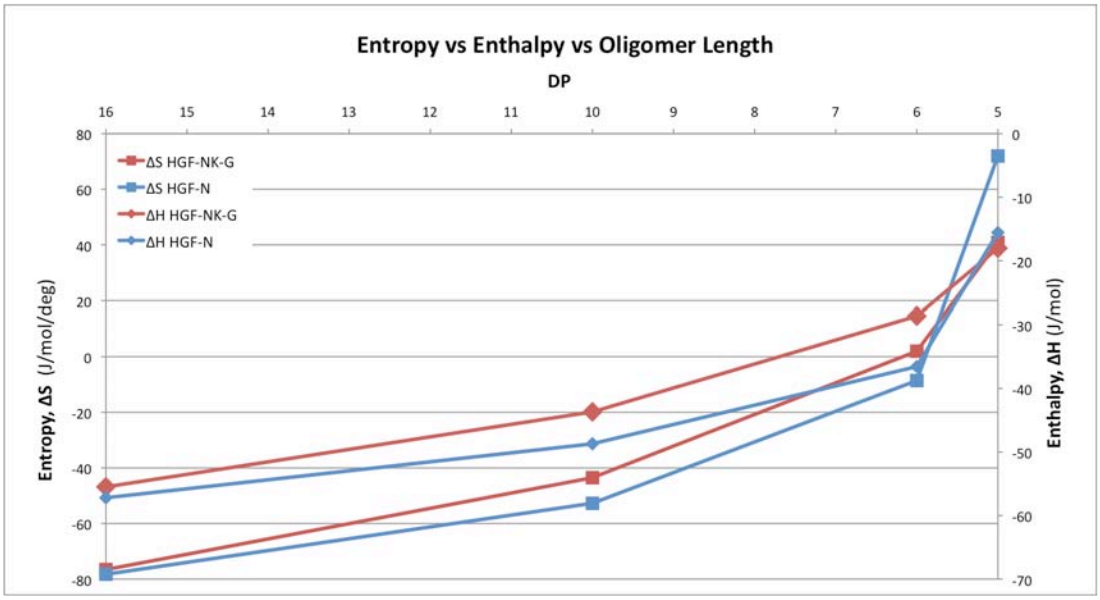


Figure 4.73 Enthalpy and entropy plotted against oligosaccharide length for HGF-N and HGF-NK-G. The error bars are too small to be visible.

A possible explanation of this trend is that the longer oligosaccharides are more unstructured than the shorter oligosaccharides and thus loose more entropy when they bind to the protein. By other words, formation of more interactions afforded by longer oligosaccharides makes the GAG more rigid, leading to losses in entropy as illustrated in Figure 4.82. It is possible that the proteins make unspecific interactions to longer heparin oligosaccharides, further contributing to an entropy loss. Also, longer oligosaccharides can occupy more extended binding sites on the proteins, or even bind two protein molecules. These factors increase the number of protein-GAG interactions, contributing to a favourable enthalpy.

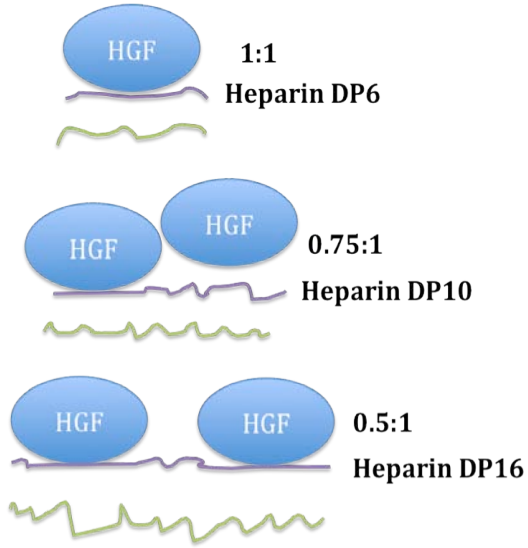


Figure 4.74 Explanation of ITC results. The green lines symbolise unbound disordered GAG with DP6 having less entropy than DP16. The purple lines symbolises a structured bound GAG that has lost entropy.

Similar results have been obtained for the interaction between extracellular superoxide dismutase and heparin [281]. In this study, it was noted that the short "oligosaccharides lose fewer degrees of freedom upon binding than the larger heparin molecules" and that this could be explained by "a large negative entropy contribution associated with more degrees of freedom being lost in heparin upon binding than for the smaller and relatively rigid oligosaccharides."

As a consequence of the enthalpy-entropy exchange observed for longer oligosaccharides, the K_D does not vary much with oligomer length (Figure 4.83). This has also been reported for the HGF-heparin interaction by Delehedde *et al* [10].

It is also evident that the stoichiometry decreases with increasing oligosaccharide length (Figure 4.73). This implies that longer heparin chains are more likely to be able to accommodate two protein molecules, suggesting that HGF-N and HGF-NK-G bind to GAGs like beads on a string. One GAG molecule binds one protein molecule

for short GAGs such as Fondaparinux and DP6, but for longer GAGs such as DP16, the ratio approaches two proteins per GAG chain. If the oligosaccharides were binding to protein molecules in a *trans* manner, two and four proteins would have been observed interacting with DP6 and DP16 oligosaccharides, respectively. This was not the case.

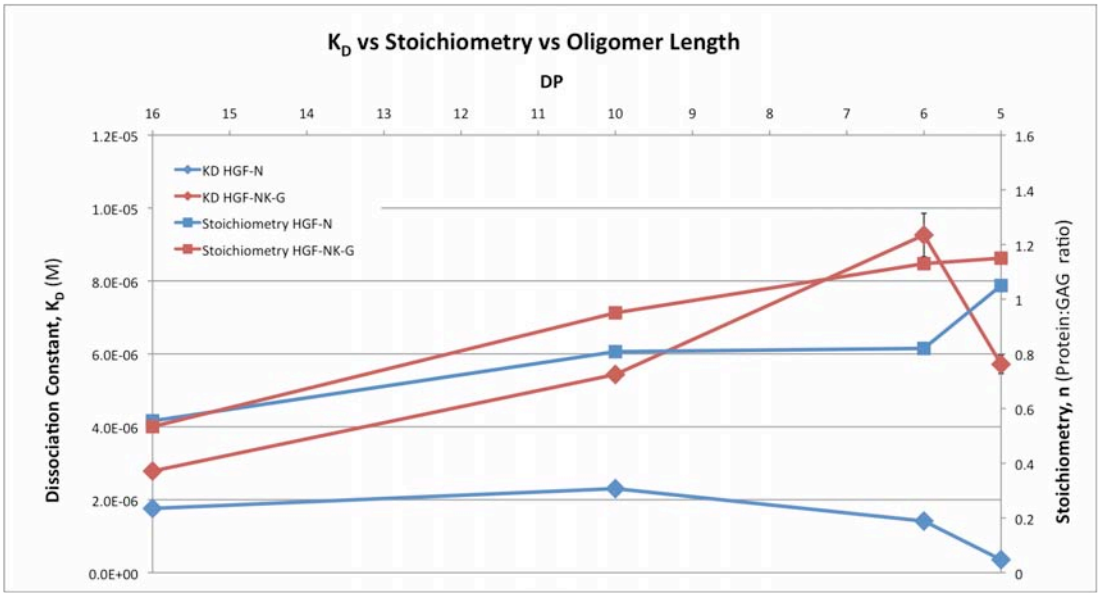


Figure 4.75 Stoichiometry and K_D data from ITC experiments plotted against heparin oligomer length. Most error bars are too small to be visible.

Overall, the ITC data presented here is in a good agreement with results previously obtained for the interaction between HGF-N and FITC-labelled LMWH heparin (average DP12), which has been shown to have a K_D of 1-2 μM and a stoichiometry of 1:0.82 [21].

The interaction of Fondaparinux with HGF-N and HGF-NK-G has a more favourable entropy than heparin hexasaccharide, which is just one monosaccharide longer. This observation can be explained by two factors: 1) The overall entropy change when a protein and a GAG form a complex depends on the loss of translational entropy and rigidifying of the GAG and protein. 2) In the opposite direction, the increased

translational entropy of the released sodium ions (the polyelectrolyte effect) and ordered water molecules covering the binding surface increases the entropy of the complex. Differences in any of these two factors could account for the more favourable entropy observed for Fondaparinux. It is possible that either the protein or Fondaparinux looses less entropy when interacting with the protein constructs than heparin hexasaccharide. This could be explained by the structural differences between Fondaparinux and heparin hexasaccharide, for example, the former contains an unsulfated iduronic acid sugar ring. By performing ITC titrations in buffers with different ionic strengths and heats of ionisation it is possible to estimate of the contribution of non-ionic interactions, deduce how many sodium ions that are released in the binding event and how many protons that are transferred [282]. It would be interesting to extend the studies presented here and to perform these experiments on Fondaparinux or other GAG mimetics and oligosaccharides.

Fondaparinux also has a less favourable enthalpy than heparin DP6, which indicates that Fondaparinux forms fewer intermolecular bonds than heparin DP6. Again, it is likely that this is due to the structural differences between Fondaparinux and heparin hexasaccharide, in particular their different sulfation pattern.

Deeper insights into GAG-protein interactions via ITC could potentially be obtained by using pure heparin species with a defined sulfation pattern. However, the formation of higher order complexes and the adaptability of the HGF – GAG interaction means that several different binding modes will likely make the atomic resolution interpretation of ITC data difficult.

4.9 NMR Monitored Thermal Denaturation

Thermal denaturation assays are common in drug discovery and ligand screening research to identify lead compounds [283]. A thermal denaturation assay is typically looking for shifts in melting temperature (T_m) upon ligand binding. Proteins are generally stabilised by ligand binding and the melting temperature increases, but the

opposite with a decreasing melting temperature can also be true [284]. Thermal denaturation monitored by NMR has an important advantage over many other thermal denaturation methods used for ligand screening: it is reporter-free. On the other hand, NMR requires relatively big amounts of material.

To my knowledge, there are no thermal denaturation studies of the N-terminal domain of HGF. To address this shortage, the melting temperatures of HGF-N in the presence and absence of heparin were determined by NMR. The results are important in that they improve the understanding of HGF-N and how stable the protein is in its free and bound state.

A series of ^{1}H 1D spectra and ^{15}N-^{1}H HSQC NMR spectra were collected at 30 °C, 50 °C, 60 °C, 70 °C, 75 °C, 80 °C, 85 °C and 90 °C in the presence and absence of ten times excess heparin decasaccharide (DP10).

The disappearance of amide signals from ^{15}N-^{1}H HSQC spectra and a clear change in 1D spectra including disappearance of methyl signals (resonating at < 1 ppm) at 60-70 °C indicated denaturation of free HGF-N as can be seen in Figure 4.76.

The NMR data suggests that HGF-N is stabilised by the presence of tenfold excess heparin DP10 and denatures at approximately 75-80 °C (Figure 4.77**)**.

In conclusion, the T_m of HGF-N is approximately 65 °C and the value increases approximately 10-15 °C in the presence of heparin.

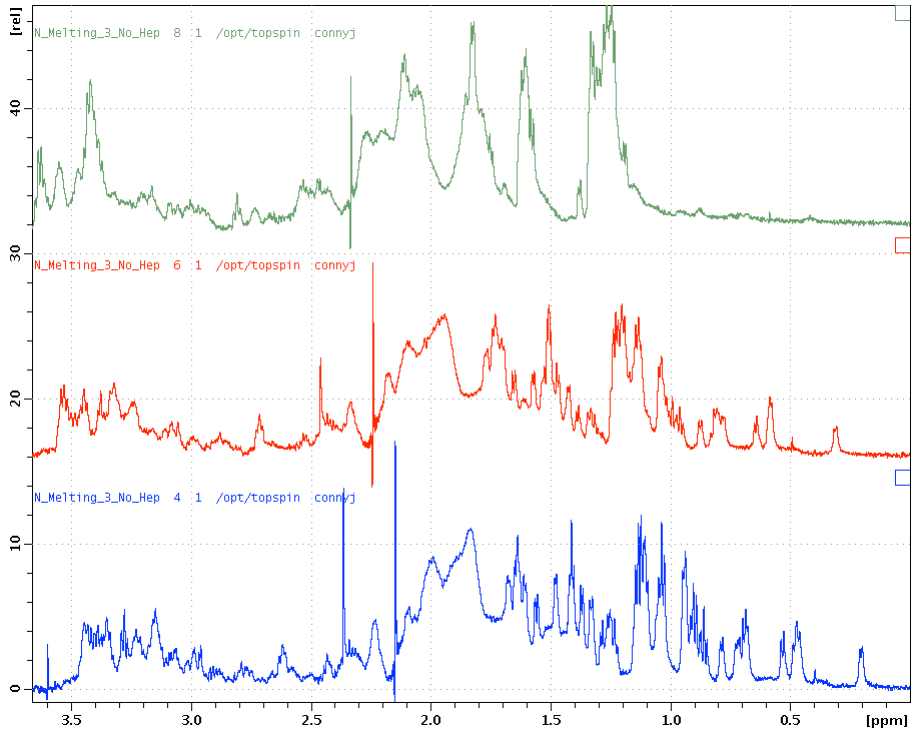


Figure 4.76 Partial ^{1}H spectra of HGF-N at the following temperatures: blue 50 °C; red 60 °C; green 70 °C.

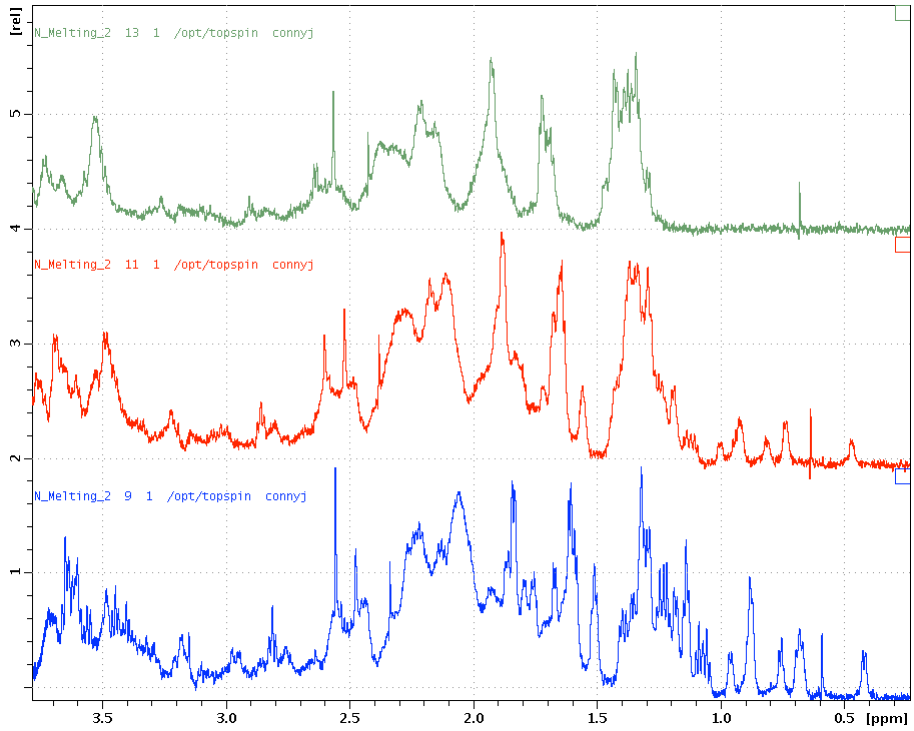


Figure 4.77 Partial ^{1}H spectra of HGF-N in the presence of ten times excess heparin at the following temperatures: blue 70 °C; red 75 °C; green 80 °C.

4.10 NMR Titrations of Crude Heparin, GAG mimetics and Enzymatically Derived DP4 and DP10 GAGs

The HGF-N – GAG interaction has previously been studied by NMR [21], [68], however, no comprehensive study comparing different GAGs and different lengths of oligosaccharides is available. In this section, the binding modes of different types of GAGs and GAG mimetics such as Fondaparinux and chemically sulfated maltose were compared and studied on a molecular level by NMR. The results show how different GAGs bind to HGF in different modes and are important in increasing the basic understanding of the HGF-GAG interaction.

By titrating different GAGs into ^{15}N-labelled HGF-N and monitoring ^{1}H-^{15}N HSQC spectra, the GAG binding site was mapped onto the protein structure and K_Ds of complexes were estimated. An example of overlaid ^{1}H-^{15}N HSQC spectra of a protein – GAG titration is shown in Figure 4.78. The movement of NH crosspeaks was quantified by combined chemical shift differences calculated using equation 2.12 as described in section 3.11, the data was then visualised in histograms as a function of residue number (Figure 4.81). The histograms were produced by comparing the titration starting point and the saturation point, which is defined as the excess of GAG where the NH crosspeaks are not moving if more GAG is added to the sample.

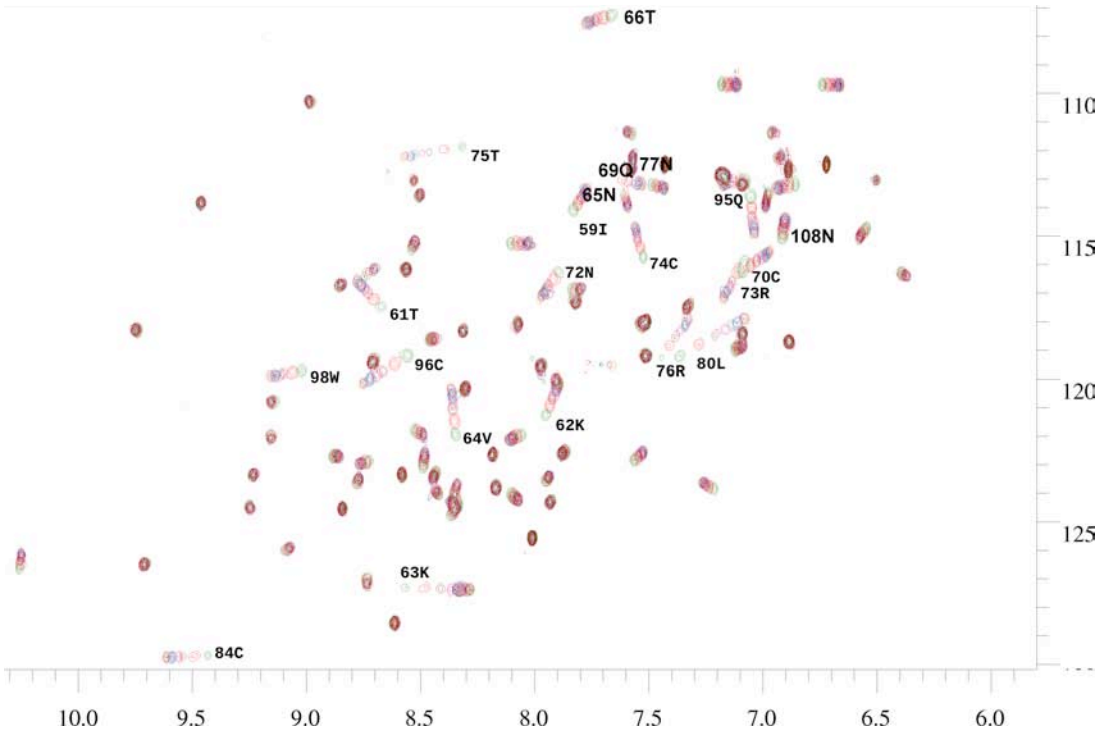
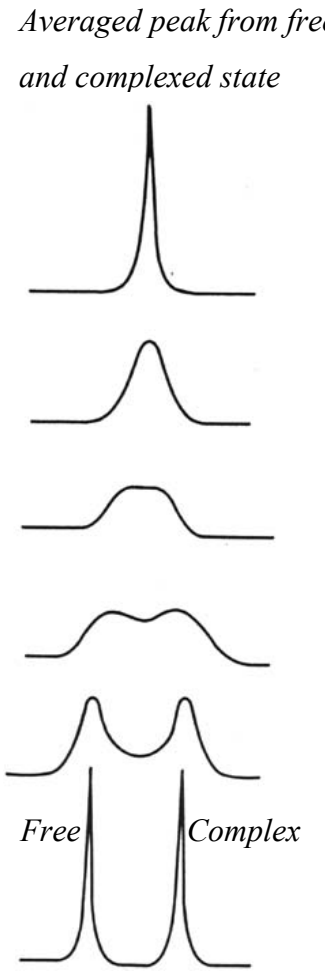


Figure 4.78 Overlaid ^{1}H-^{15}N HSQC spectra HGF-N-Δ-6 in the presence of fully sulfated heparin-derived tetrasaccharide (DP4C). The protein:ligand ratio was varied between 1:0.1 to 1:6.

Some NH cross peaks are not visible in ^{1}H-^{15}N HSQC spectra of HGF-N due to protein dynamics on an unfavourable timescale. The crosspeaks that cannot be seen in uncomplexed ^{1}H-^{15}N HSQC spectra are: Gly31-Asn37, Asn77, Lys78 (weak), Gly79, Ser106 (very weak), Tyr107 (very weak).

The majority of the visible HGF-N ^{1}H-^{15}N HSQC crosspeaks were found to exhibit fast-exchange attributes when titrated with ligand: the crosspeaks were moving stepwise as ligand was titrated. Some crosspeaks showed signs of intermediate exchange: they initially weakened but sharpened as more of the ligand was titrated in. A peak in intermediate exchange might have distorted shape and will appear weaker and broadened (Figure 4.79).



- **Fast exchange:** weak affinity ligand, fast k_{off}.
Peak shape is getting narrower as the free state and complexed states interchange so quickly that they are not observable on an NMR-timescale.

- **Intermediate exchange:** medium affinity ligand, medium k_{off}.
Peak shape is getting broader as both the free and complex peak contribute to the peak shape. The protein spends significant time in both complexed state and free state on an NMR time-scale.

- **Slow exchange:** high affinity ligand, slow k_{off}.
The on/off rate of the ligand is so slow on an NMR time-scale that two distinct peaks are observed in the spectrum.

Figure 4.79 Peak shapes of slow, intermediate and fast exchange NMR peaks of a protein – ligand complex.

The higher affinity interaction, the more time the ligand will be part of the complex and the off-rate (k_{off}) will be slower. High affinity interactions are therefore more likely to give rise to slow exchange spectra when a ligand is titrated. As a rule of thumb, interactions with a K_D of less than 10 μM fall into a slow or intermediate exchange. However there are many exceptions to this general rule: slow exchange has been shown for the interaction between Hsp70 and a peptide with a K_D of 500 μM (because of slow k_{on}) and fast exchange was encountered for the binding of a phosphate compound to haemoglobin with a K_D of approximately 1 nM (because of a multisite binding mechanism) [285].

Several protein-GAG interactions with K_Ds in the μM range have been characterised by NMR and shown to exhibit fast exchange attributes when the protein is titrated with ligand. The 13-14F3 fragment of fibronectin and its interaction with heparin has been shown to have a K_D of 2.9 μM and exhibit fast exchange NMR attributes [286]. The N and C -terminal domains of Pleiotrophin bind heparin DP14 with K_D values of 53 μM and 20 μM, respectively, and show fast exchange properties [287]. The SOS - FGF-1 interaction has been shown to have a K_D of 59 μM and the interaction between the ligand binding D2 domain of FGFR binds SOS with a K_D of 27 μM and both interactions show fast exchange attributes [288].

The interactions between HGF-N, heparin and SOS have previously been shown to exhibit typical fast exchange properties [21]. The interaction between HGF-N and FITC-labelled LMWH heparin (average DP12) has been shown to have a K_D of 1-2 μM and a stoichiometry of 1:0.82 [21]. These findings are in excellent agreement with the results from the NMR and ITC experiments in this thesis.

The peaks that exhibited signs of an intermediate exchange when titrated with heparin oligosaccharides and Fondaparinux were typically the peaks that displayed large CCSDs, in particular Arg73 and Arg76. This is easily explained, as the classification of signals into different exchange regimes depends on the relationship between k_{off} and the NH chemical shift difference of the free and bound protein. Peaks with larger chemical shift separation therefore are more likely to be in an intermediate exchange regime. Because of this, the K_D of the Fondaparinux - HGF-N interaction is not reported in section 4.10.4.

4.10.1 Crude Heparin

It has been reported that HGF-N precipitates upon addition of crude heparin [17]. However, as results often differ between heparin batches, conditions and laboratories, it was attempted to titrate crude bovine lung heparin into ^{15}N-labelled HGF-N, monitoring ^{15}N-^{1}H HSQC spectra. When heparin was added to the sample, it

instantaneously went cloudy, but as the protein did not precipitate and sink to the bottom, it is believed that large complexes are formed that stay in solution but are too big to be observed by NMR (Figure 4.80). Some crosspeaks come back when a very large excess of heparin is added, these crosspeaks mostly belong to side chains and termini.

When a large excess of crude heparin is added to an NMR sample, the salt concentration soon becomes problematic. For example, if heparin is added to fiftyfold excess, every heparin chain contains fifty Na^+ atoms and the protein concentration is 100 μM, the increased Na^+ concentration is 250 mM, which severely decreases the sensitivity of the experiment, especially when using instruments equipped with a cryoprobe. To compensate for this effect, the number of scans was increased 2.5 times for the 60:1 spectrum compared to the other spectra (Figure 4.80).

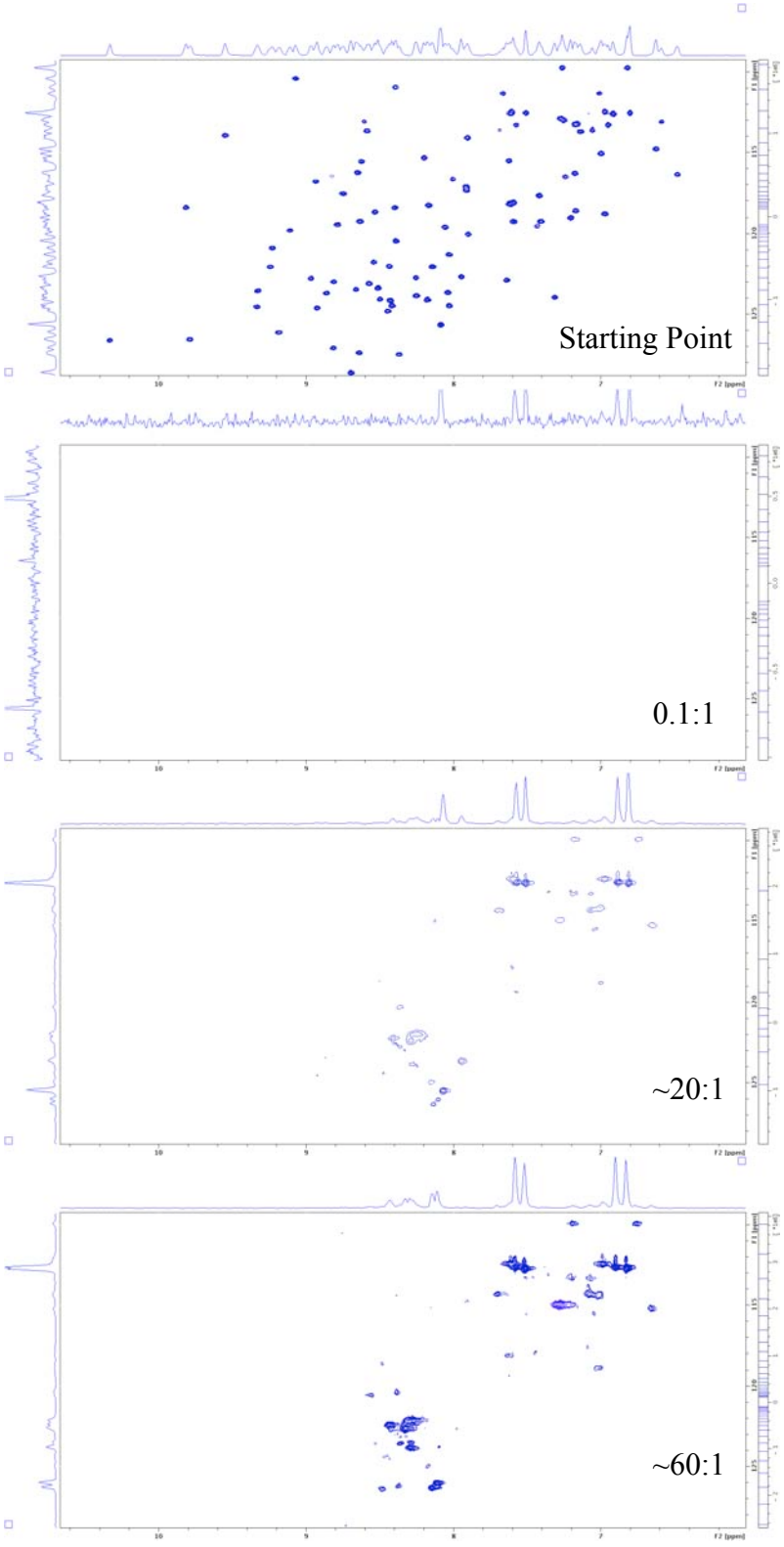


Figure 4.80 Titration of crude heparin into ^{15}N-labelled HGF-N monitored by ^{15}N-^{1}H HSQC spectra. The numbers indicate GAG:Protein ratio.

4.10.1 Fully Sulfated Heparin DP4 and Fondaparinux

Two protein constructs were used for Fondaparinux and heparin NMR titration experiments: HGF-N and HGF-N-Δ6. The later construct does not contain the first six N-terminal residues (GQRKRR). At physiological pH, the arginines and lysine in this sequence are positively charged and could therefore be involved in GAG binding. This suggestion is supported by the fact that HGF-N-Δ6 was found to bind weaker to a heparin column than the HGF-N (section 4.5.2) and by previous studies [63]. Fortuitous cleavage of HGF-N by a protease during secretion of *P. pastoris* (section 4.2.1) allowed the contribution of these N-terminal residues to heparin binding to be studied by NMR. These N-terminal residues are flexible and neither X-ray nor NMR can determine their positions, consequently their role in GAG binding must be studied indirectly.

It is worth pointing out that a high density of positive charge does not automatically translate into heparin affinity. For example, module 13 of factor H has the highest pI value among the twenty modules of factor H, yet it does not bind heparin [289]. HGF-N is a 127-residue protein that contains 16 lysines and 7 arginines; out of these four lysine and two arginine backbone NH resonances show significant combined chemical shift differences (CCSDs) during GAG titrations (Figure 4.81). The non-moving basic residues are lysines 43, 44, 47, 52 of the β1 strand and lysines 109, 110 and 122 of the β5 strand that create a positive patch on one face of HGF-N (Figure 4.82).

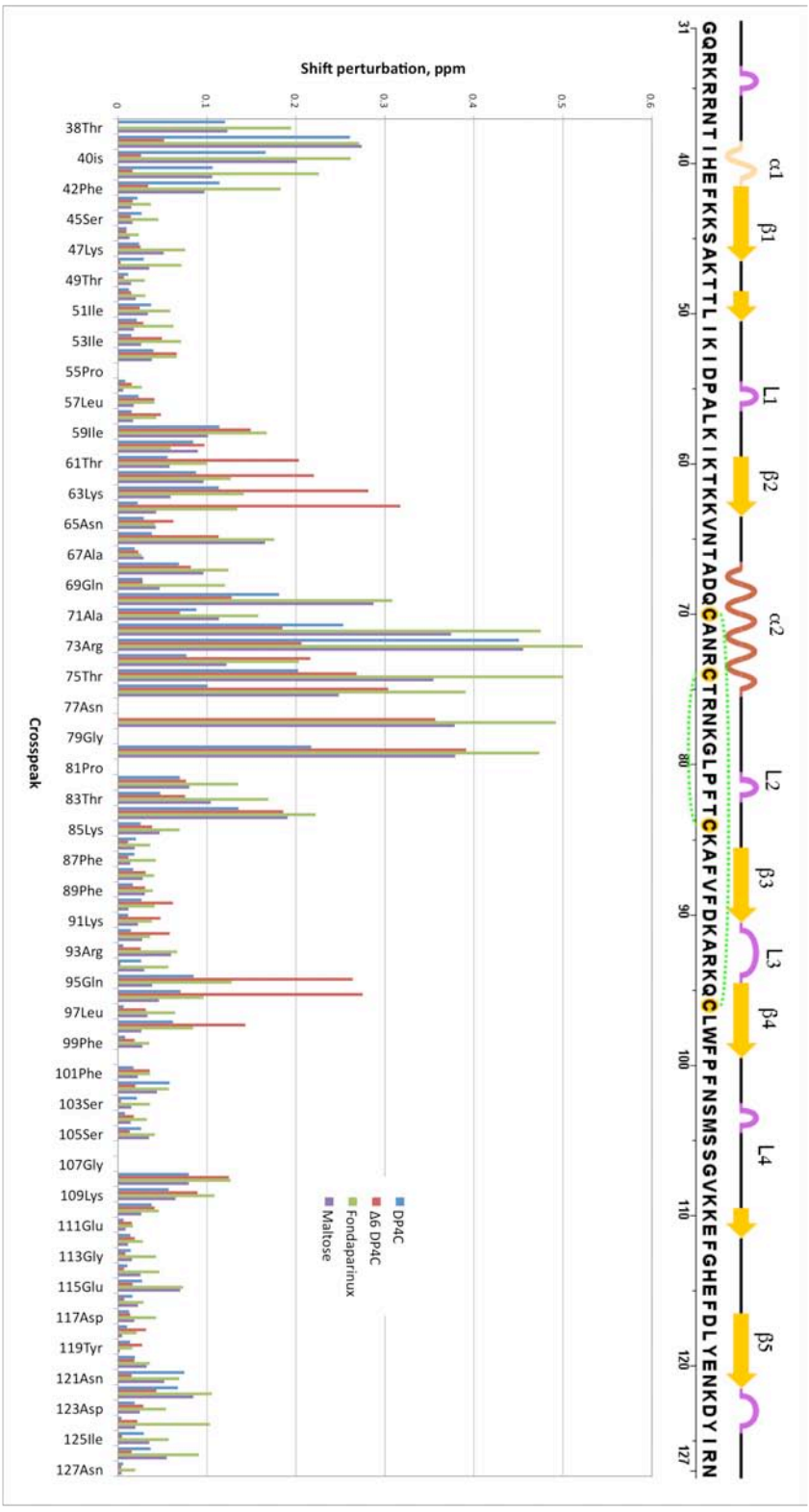


Figure 4.81 Histogram comparing different oligosaccharides titrated into HGF-N monitored by ^{1}H-^{15}N HSQC spectra. Combined chemical shift perturbation is displayed on the y-axis and residue number on the x-axis. Δ6: HGF-N-Δ6 was used instead of HGF-N. Maltose: Chemically sulfated maltose.

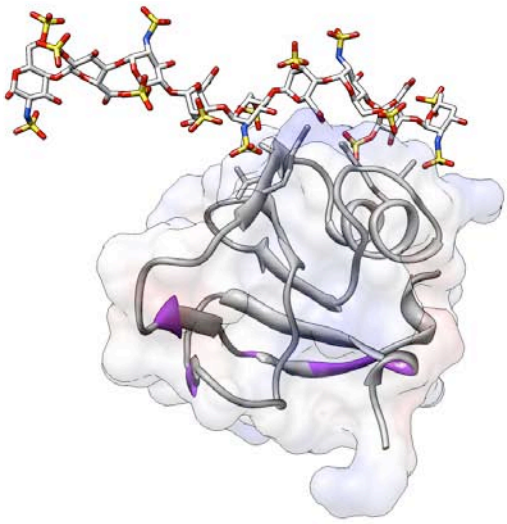


Figure 4.82 A protomer from the 1GMO PDB file was rendered in UCSF Chimera with transparent columbic surface colouring, showing residues 43, 44, 47, 52, 109, 110, and 122 on the ribbon structure in purple.

Titrating fully sulfated heparin tetrasaccharide (DP4C) into HGF-N caused large combined chemical shift differences for residues 70-84, which are part of the α2-helix and the L2 loop. The second most affected region of the protein is comprised of residues 38-42, where T38 is the first residue that is visible in ^{1}H-^{15}N HSQC spectra of HGF-N. None of the NH resonances of the first six residues are observable by NMR due to the movement of these residues on an unfavourable timescale [21], [17]. Residues 38-42 (TIHEF) are not positively charged, and although they can play a supportive role in the GAG binding through hydrogen bonding, it is likely that the basic four N-terminal amino acids (RKRR) are more important. Secondary effects from the more N-terminal residues therefore likely cause the CCSDs of residues 38-42. The first indirect evidence for this hypothesis comes from the appearance of new NH crosspeaks in ^{1}H-^{15}N HSQC spectra of HGF-N during the course of titration (Figure 4.83). These new signals which move during the titration likely belong to the 'invisible' N-terminal residues, which become rigidified upon GAG binding [21]. As expected, most of these new crosspeaks cannot be seen when titrating heparin DP10 into HGF-N-Δ6.

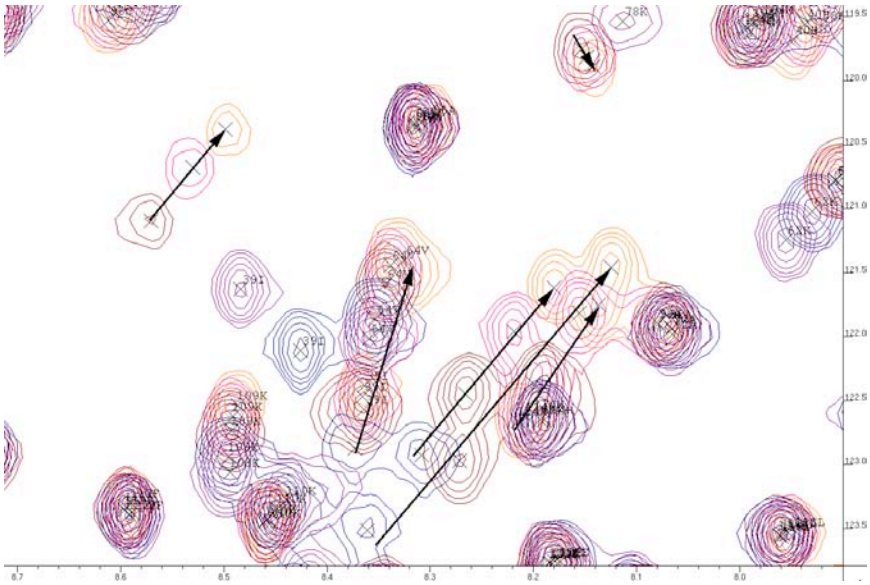


Figure 4.83 The appearance of new peaks as Fondaparinux is titrated into HGF-N. The purple spectrum is the starting point, blue: 0.5X excess Fondaparinux over the protein, brown: 2X excess, pink: 6X excess, orange: 20X excess.

The appearance of new NH crosspeaks is more evident when titrating Fondaparinux compared to when titrating heparin DP4C. This is likely due to the extra saccharide unit and 3-O-sulfate of Fondaparinux.

The second line of evidence for the involvement of the N-terminus in GAG binding comes from comparing the CCSDs caused by titrating DP4C into HGF-N and HGF-N-Δ6 (Figure 4.81):

a) For HGF-N, residues 38-42 are affected, but for HGF-N-Δ6, these residues are not affected. This implies that by loosing the positively charged N-terminus, the binding to this part of the protein is abolished.

b) For the more native HGF-N protein construct, residues 70-73 of the α2-helix are affected more than the rest of the helix and the following loop. For HGF-N-Δ6, residues 74-80 (L2) are affected more than 70-73 (α2). This implies that by loosing the positively charged N-terminus, the binding mode of DP4C changes.

c) Only HGF-N-Δ6 shows large CCSDs for the β2 strand residues 59-64+66 and β4 strand residues 95-96. This again implies that the binding mode of DP4C is different for HGF-N-Δ6 compared to HGF-N. The differences are summarised in Figure 4.84.

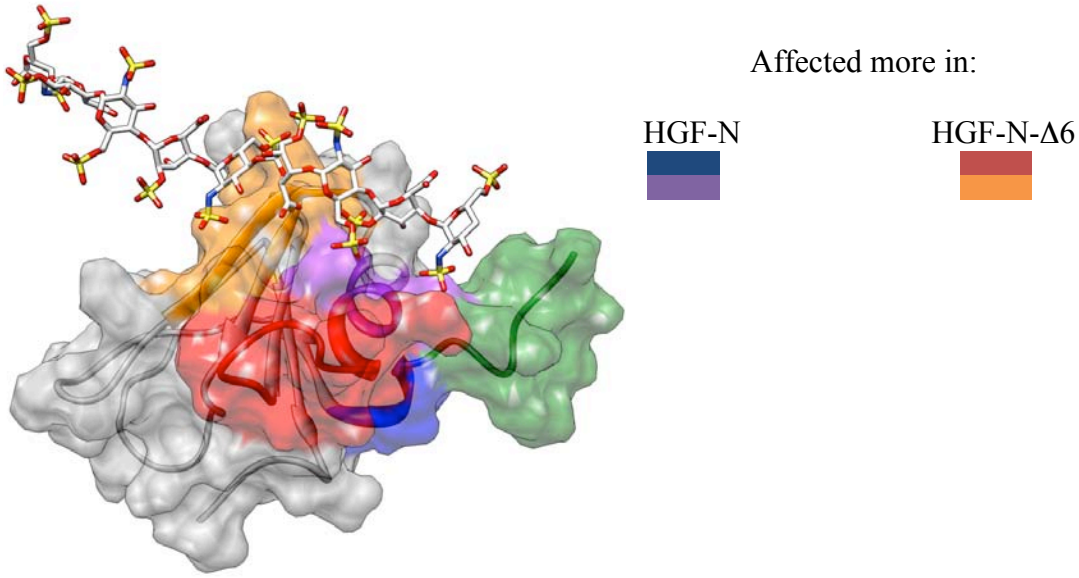


Figure 4.84 The 2HGF PDB file was rendered in UCSF Chimera and overlaid with a heparin oligosaccharide from the 1GMO PDB file. Green: residues 31-37. Blue: 38-42. Purple: 70-73. Red: 74-84. Orange: 59-64, 66, 95 and 96.

The results imply that when titrating shorter heparin oligosaccharides into HGF-N, the oligosaccharides bind preferably to residues 70-84 of the α2-helix and residues 38-42 of the N-terminal region that is disrupted in HGF-NK-G and HGF-N-Δ6. This native HGF-N binding mode is different from the binding mode of HGF-N-Δ6 and HGF-NK-G in the heparin co-crystal structure where residues 59-64, 66, 95 and 96 of the β2/4-strands are more important because the oligosaccharide has moved away from the disrupted N-terminus. Therefore, it would be interesting to dock the heparin to HGF-N with HADDOCK (http://www.nmr.chem.uu.nl/haddock/), AutoDock (http://autodock.scripps.edu/) or similar docking software to further clarify the differences.

To summarise, the HGF-N heparin binding site consists of three binding regions: the main α2-helix binding region (residues 70-84), the N-terminal region (residues 32-42) and the β2/4-strands region (residues 59-64, 66, 95 and 96). For HGF-N, the main α2-helix binding region and the N-terminal region are the most important binding regions. For HGF-N-Δ6, the main α2-helix binding region and the β2/4-strands region are the most important binding regions. Only one tetrasaccharide is bound at a given time to the protein. The results are illustrated in Figure 4.85.

a) b)

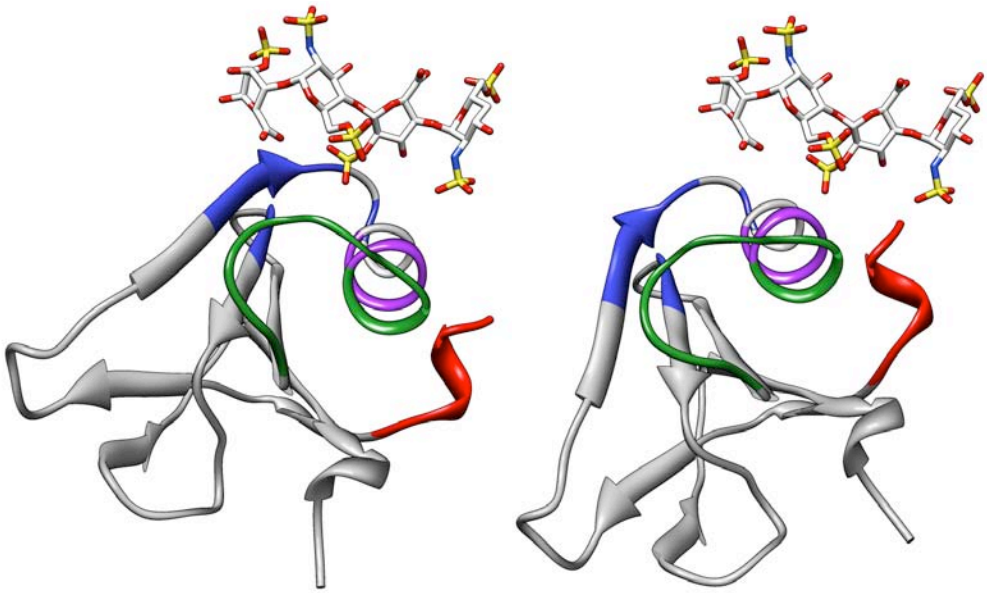


Figure 4.85 a) The N domain of a protomer from the 1GMO PDB files was rendered in UCSF Chimera with only four sugar units displayed and residues 38-42 (N-terminal binding region) coloured in red, residues 70-73 coloured in purple, residues 74-84 coloured in green (α2-helix main binding region) and residues 59-64, 66, 95 and 96 (β2/4 binding region) coloured in blue. a) This binding mode is seen for HGF-NK-G and HGF-N-Δ6. b) The protein has been manually rotated in relation to the heparin oligosaccharide to put the N-terminus closer to the GAG, visualising the different binding mode of the more native HGF-N.

To further confirm that only one DP4C molecule binds to native HGF-N, the stoichiometry of DP4C binding to HGF-N was studied by MALDI-TOF. DP4C was titrated from 0.1 to 20–fold molar excess. The MALDI-TOF spectra are characteristic of a 1:1 protein:ligand interaction as only one higher MW species can be seen as the DP4C ligand is titrated (Figure 4.86).

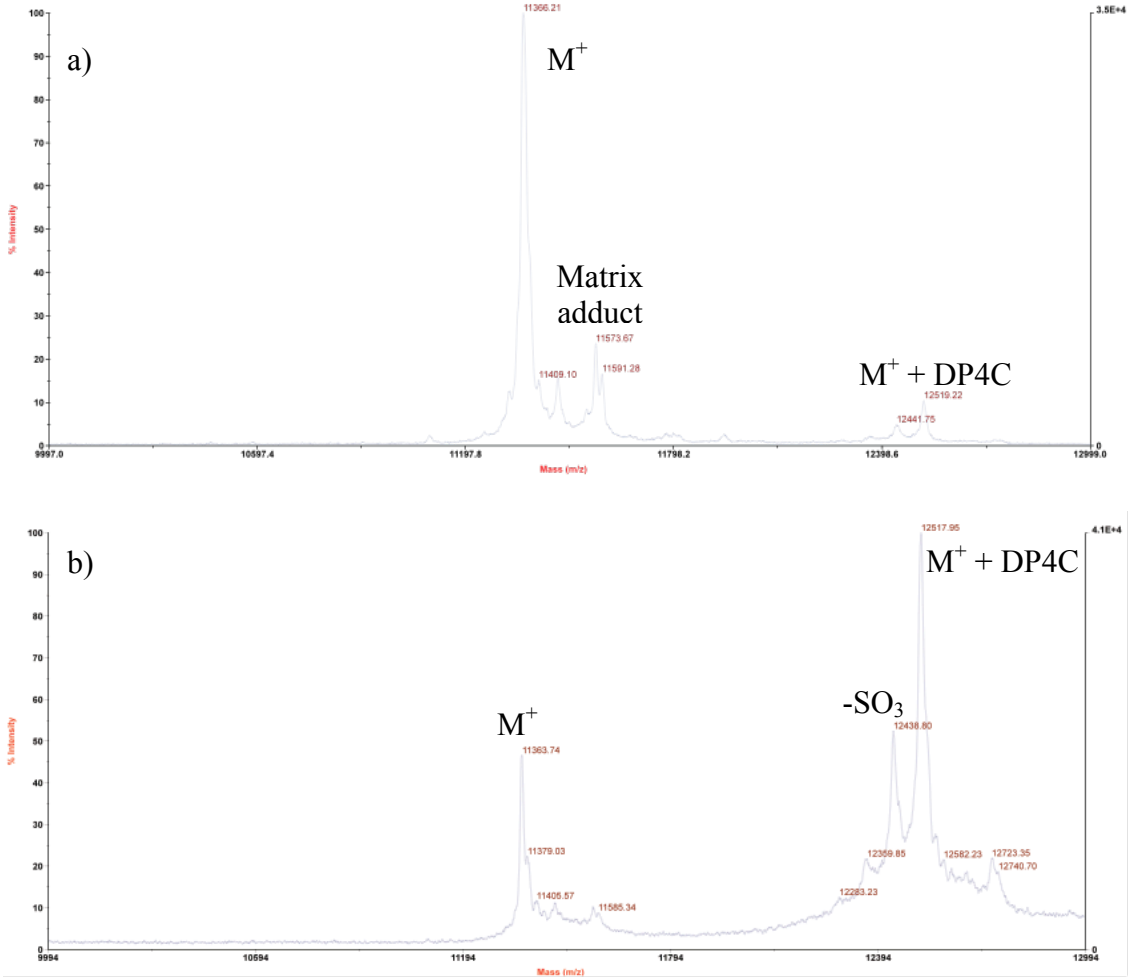


Figure 4.86 MALDI-TOF titration of DP4C into HGF-N. a) 0.5:1 DP4C:HGF-N ratio. b) 20:1 DP4C:HGF-N ratio.

The pentasaccharide Fondaparinux was titrated into HGF-N. Compared to DP4C, it is one monosaccharide unit longer, it does not contain a modified uronic ring with a double bond from enzymatic cleavage, it has one 3-O-sulfate and it is a Me-

glycoside (Figure 1.22). The CCSDs upon titration of HGF-N with Fondaparinux are very similar to those observed for HGF-N titrated with DP4C (Figure 4.81). The only major difference is that Fondaparinux causes larger changes for the N-terminal and main α2-helix binding regions. This could be due to the increased length, extra sulfate and that it is a Me-glycoside.

When plotting the residues with a combined chemical shift perturbation >0.12 and >0.17 ppm for Fondaparinux, i.e. the residues that are most affected by binding on the structure of HGF-N, a single well defined binding site that matches well with earlier mutational, NMR and crystallographic studies can be seen [14], [21] (Figure 4.87).

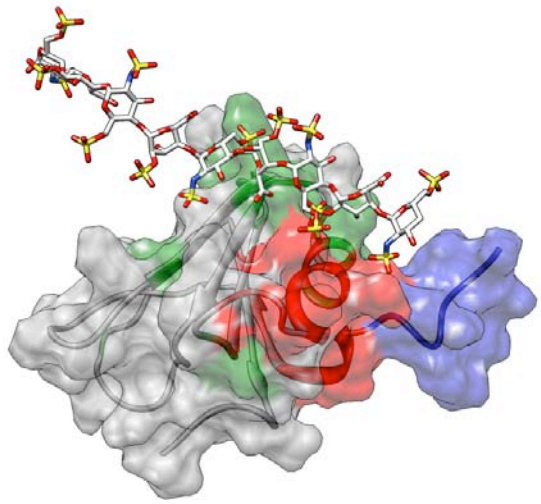


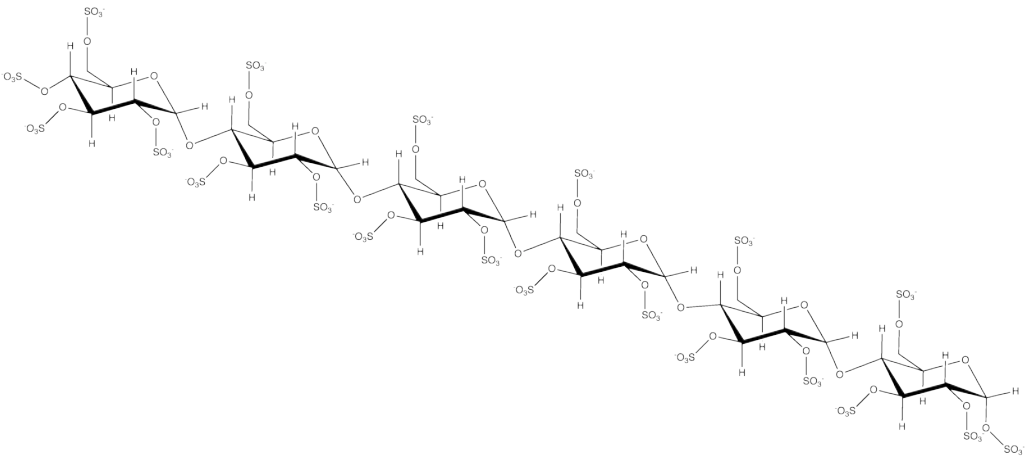
Figure 4.87 The 2HGF PDB file was rendered in UCSF with a heparin oligosaccharide from the 1GMO PDB file overlaid. Residues 31-37 are coloured in blue, residues with a CCSD > 0.17 ppm are coloured in red, residues with a 0.12< CCSD < 0.17 ppm are coloured in green.

4.10.2 Chemically Sulfated Maltose Oligosaccharides

Chemically O-sulfated maltose is a glucose-based disaccharide with a D-Glucopyranosyl-D-glucose structure that is similar to sucrose octasulfate, a widely used GAG mimetic (Figure 4.88). Chemically O-sulfated maltohexaose is a GAG

mimetic similar to Muparfostat (section 1.2.6). In an attempt to determine the degree of chemical O-sulfation of the species used in these experiments, MALDI-TOF spectra of chemically sulfated maltohexaose were collected. The spectra showed extensive loss of $NaSO_3$, as often seen for similar species [290], which made it difficult to clarify if the sample was homogeneously sulfated or not. The spectra showed a range of peaks, corresponding to loss of all sulfates except one or two, up to fully sulfated maltohexaose (data not shown). It is therefore unknown what the sulfation degree was as it cannot be said if the sulfates were lost during MALDI-ionisation or if the compound was not fully sulfated.

a)



b)

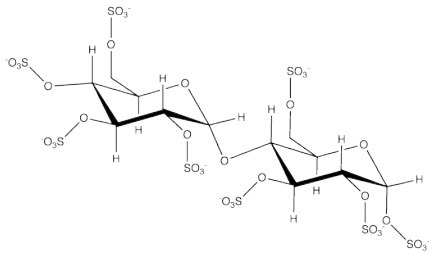


Figure 4.88 a) Structure of chemically sulfated maltohexaose. b) Structure of chemically sulfated maltose.

Initially, titration of maltohexaose into HGF-N at 0.1 to 0.4 molar excess caused CCSDs similar to other GAGs, but as the maltohexaose was added, the protein precipitated. In the ^{1}H spectra, protein peaks decrease in intensity and the sugar

peaks increase in intensity (Figure 4.89), indicative of that the protein precipitates and that the oligosaccharide remains in solution and no longer binds to the protein. It is likely that a combination of the length and very high degree of sulfation of the carbohydrate are the main reasons for causing protein precipitation.

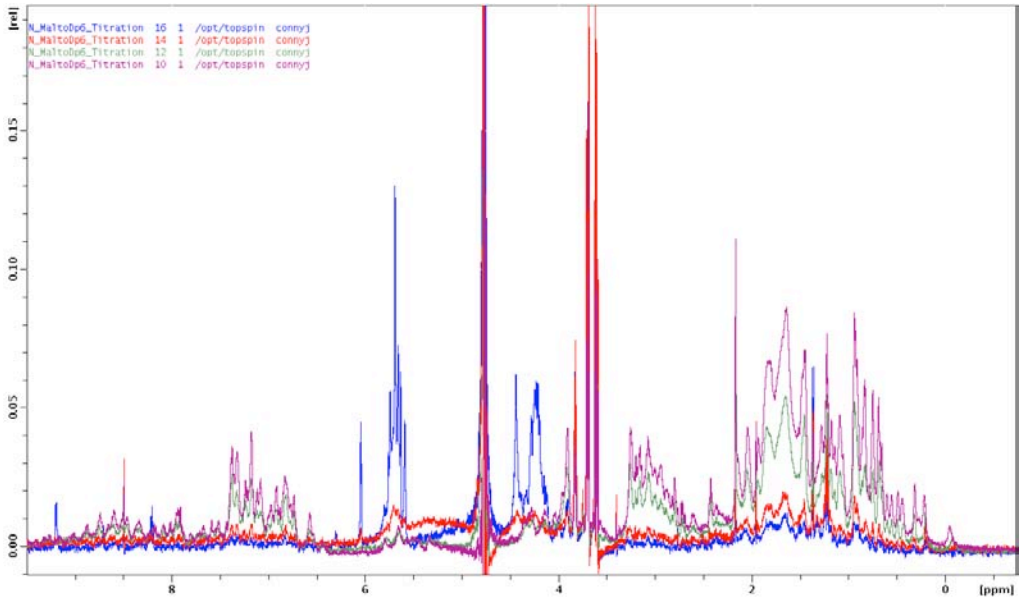


Figure 4.89 ^{1}H NMR spectra of chemically sulfated maltohexaose titrated into HGF-N at the following maltohexaose : protein ratios; purple: 0.4:1, green: 0.8:1, red: 1.6:1, blue: 5:1.

Chemically sulfated maltose did not precipitate the protein and was used instead. This compound is similar to sucrose octasulfate (SOS), a GAG mimetic that is often used in experiments when crude or digested heparin is not suitable. There are examples in the literature where heparin and SOS have been binding to the same binding site and in similar modes to a protein despite the different structures [21], [291].

The sulfated maltose bound to the main α2-helix and N-terminal regions, but induced small CCSDs for the β2/4 region, producing a CCSD profile similar to that for Fondaparinux and DP4C (Figure 4.81).

4.10.3 Heparin, CSC, CSA and DS Decasaccharides

After titrating relatively short oligosaccharides, longer GAGs were titrated to better mimic the native *in vivo* GAG polysaccharides. When comparing these results with those obtained for shorter oligosaccharides, it is clear that the overall trends are similar, but some interesting differences are apparent.

Interestingly, for the Fondaparinux, sulfated maltose and DP4C titrations, differences between HGF-N and HGF-N-Δ6 can be seen in the N-terminal binding region, the main α2-helix binding region and the β2/4-strands binding region (Figure 4.81). However, when titrating heparin decasaccharide, differences between the two proteins can only be seen in the N-terminal region where the CCSDs for HGF-N-Δ6 were much smaller (Figure 4.90). This is however expected, as a decasaccharide is long enough to span the N-terminal, α2-helix and β2/4-strands binding regions, differences should only be seen in the N-terminal binding region that is lacking in the HGF-N-Δ6 protein.

For a long GAG to be able to bind across the whole binding surface, utilising all binding regions, it has to be able to adopt an optimal conformation to maximise intermolecular bonds. This explains HGF's preference for flexible IdoA sugar units [67] and the interaction could possibly be aided by a kink in the GAG, just like for FGF [194]. HGF-N is also itself likely to undergo conformational change and possibly, the flexible N-terminus rearranges, reaches towards the GAG and rigidifies as discussed above and evidenced by the appearance of new crosspeaks in ^{1}H-^{15}N HSQC spectra of HGF-N titrated with Fondaparinux (Figure 4.83). Similarly, five new ^{1}H-^{15}N HSQC crosspeaks can be seen when titrating heparin decasaccharide into HGF-N. On the other hand, only one new crosspeak can be seen for HGF-N-Δ6 titrated with heparin decasaccharide (data not shown).

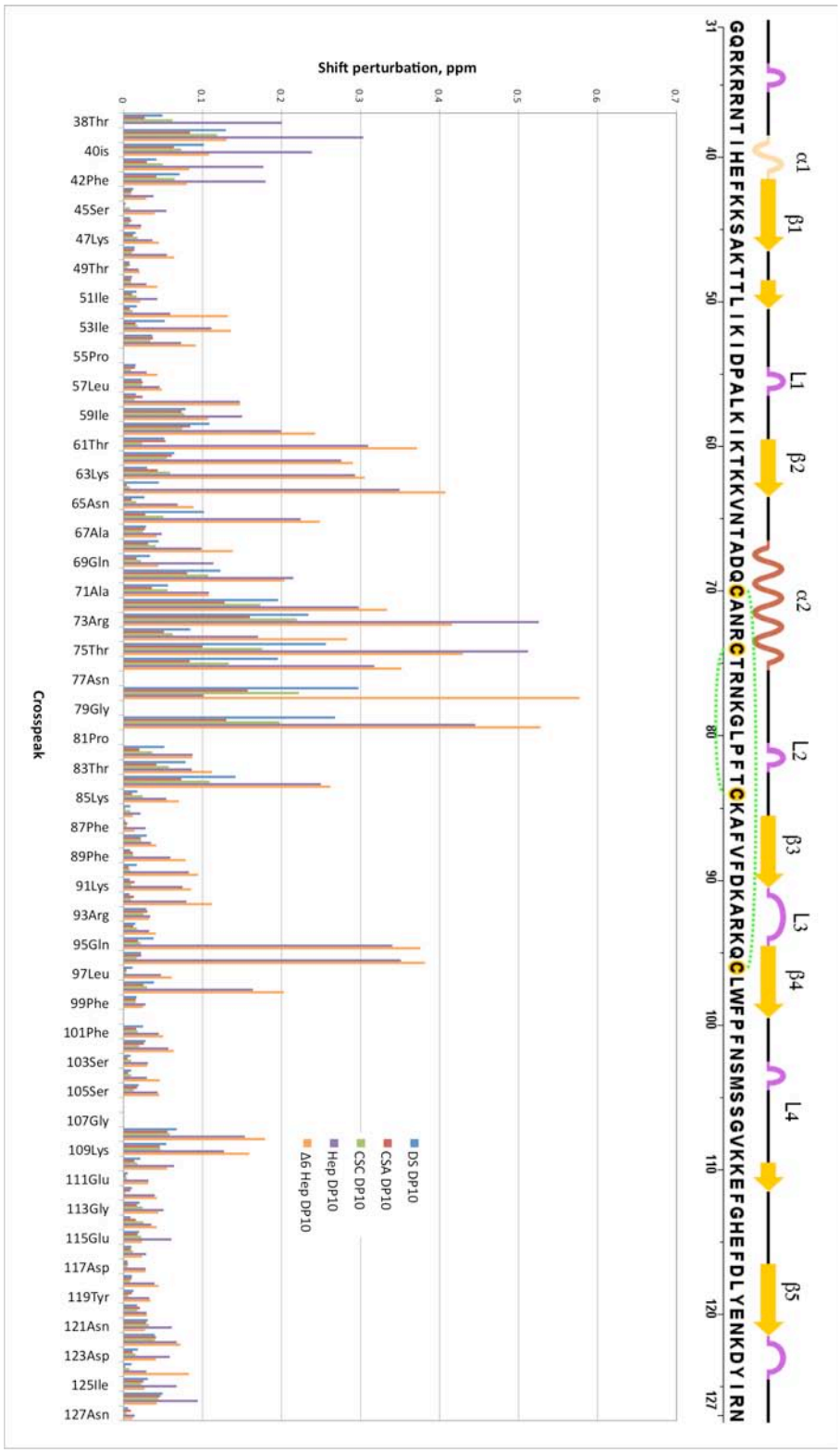


Figure 4.90 Histogram comparing different decasaccharides titrated into HGF-N monitored by ^{1}H-^{15}N HSQC spectra. Combined chemical shift perturbation is displayed on the y-axis and residue number on the x-axis.

Interestingly, it seems like heparin binds in a different mode compared to CSC, CSA and DS. Despite the decasaccharides being relatively long and able to span the whole binding site, the CS and DS oligosaccharides only induce very small CCSDs in the N-terminal and β2/4-strands region, but relatively large CCSDs in the main α2-helix region. Heparin induces shift changes in all three regions and also induces larger CCSDs than the other GAGs (Figure 4.90). CSA, CSC and DS have lower sulfate densities than heparin: CS/DS typically have one sulfate per disaccharide whereas heparin typically has three sulfates per disaccharide. This probably explains the smaller shift differences for CS and DS and different binding modes of the different GAGs. CSA, CSC and DS exhibit similar CCSD profiles and thus bind in similar modes despite the different types of sugar unit backbones and sulfation patterns. Overall, the relative order of the CCSDs is heparin > DS > CSC > CSA, however, the differences in CCSDs do not translate into differences in K_D as discussed in section 4.10.4. The inability of CSA, CSC and DS to utilise the whole binding site and their lower CCSDs explain their lower affinity for HGF-N compared to heparin. The results from all protein – oligosaccharide titrations are summarised in Figure 4.91.

In a previous study, Deakin *et al* conducted NMR titrations of heparin tetrasaccharide and DS hexasaccharide into ^{15}N labelled HGF-NK-G [68]. Although clear differences between the two oligosaccharides were seen in this study, the overall binding site and binding mode was thought to be the same and it was hypothesised that the differences were possibly due to differences in lengths and structures. Even if two different GAGs induce somewhat different CCSD profiles, it is likely that they can bind in similar modes to the same binding site.

	N-terminal	Main α2-helix	β2/4-strands
Hep DP4C	+	+	-
Hep DP4C HGF-N-Δ6	-	+	+
Fondaparinux	+	+	-
Sulfated Maltose	+	+	-
Hep DP10 HGF-N-Δ6	-	+	+
Hep DP10	+	+	+
CSA DP10	-	+	-
CSC DP10	-	+	-
DS DP10	-	+	-

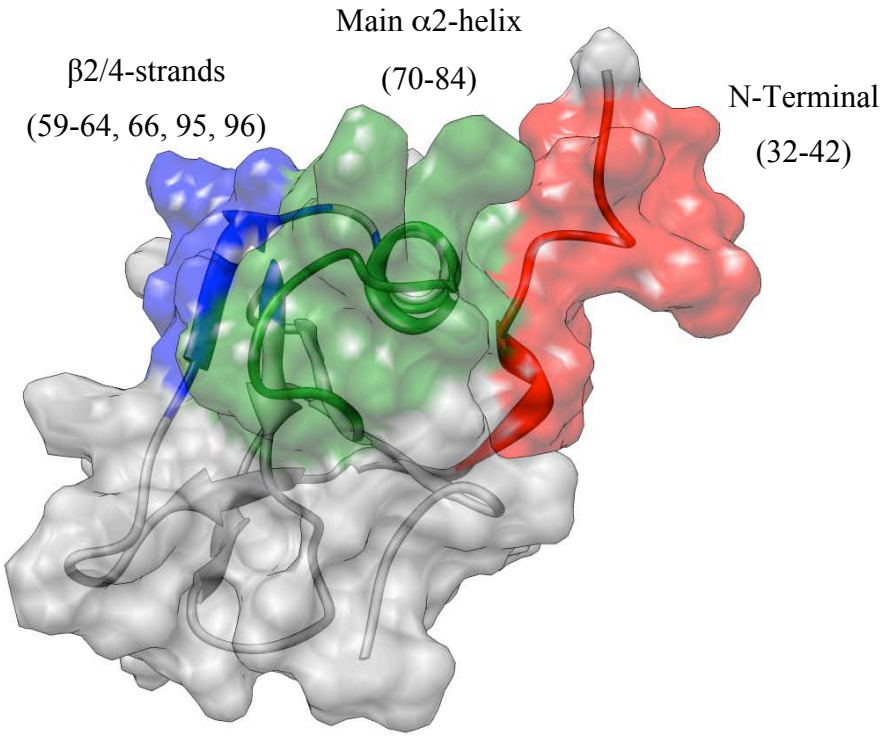


Figure 4.91 The rendered picture of HGF-N shows the three most important GAG binding regions within the protein binding site. The table summarises the results from the titrations of different types of GAGs.

4.10.4 K_D Determination

NMR titration data can be used to quantify the strength of the studied interaction. There are several NMR approaches for determining dissociation constants for protein-ligand interactions. A common approach that was employed in this project in conjunction with binding site mapping is to analyse a series of ^{1}H-^{15}N HSQC spectra as a ligand is titrated into the protein [292-297].

For the K_D determination to be straightforward and reliable it is important that the interaction falls into a fast exchange regime and the chemical shifts and relaxation rates are completely averaged between the bound and free states. If this is true, the gradual chemical shift change corresponds to the fraction of bound ligand and it is possible to use the CCSDs of NH crosspeaks as they move from the free to the bound state to determine the K_D. As a general rule, only the residues that exhibit the largest chemical shift changes and are in the fast exchange regime should be used. These usually coincide with the main residues of the binding site. In the case of protein-GAG interactions, many residues can be involved as GAGs usually bind over a large surface rather than in a small pocket. The precision of the method is limited and typically a spread of K_D values will be obtained for different residues.

Furthermore, monitoring the backbone NH chemical shift change might not be a good reporter for binding as it is often the side chains of the amino acids that are involved in binding. NMR and MS techniques to overcome this problem have recently been developed [298].

To obtain the K_D values, either a global fit in a scientific plotting program such as GraphPad Prism (http://www.graphpad.com/prism/) or Erithacus Grafit (http://www.erithacus.com/grafit/) can be employed. An alternative approach that was employed in this study is to take the average of K_D determined for all the main binding residues. Recently, a program called Auto-Face has been developed that can

automatically produce a CCSD histogram and calculate the K_Ds for many different types of interactions [293].

A 1:1 interaction was assumed for all interactions in these experiments and the K_D values were determined as described in section 3.11. A typical titration curve is shown in Figure 4.92.

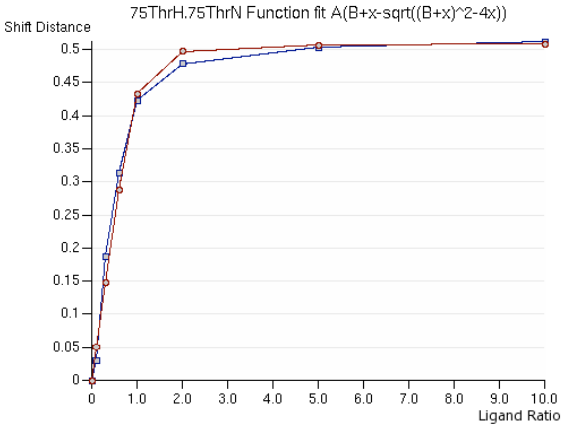


Figure 4.92 Typical ^{1}H-^{15}N HSQC titration fitting curve from the CcpNmr Analysis software. Red: fitted curve, blue: experimental curve.

Table 4.7 summarises the determined K_Ds for the interactions between different GAGs, sulfated maltose, HGF-N and HGF-N-Δ6. The K_Ds vary between 11.6 and 84.6 μM.

It was expected that the HGF-N-Δ6 protein would display lower affinity for heparin DP4C than native HGF-N as suggested by heparin affinity chromatography. However, practically identical values were obtained. It is noticeable that the error for the titration using HGF-N-Δ6 protein is lower and a general observation is that the ^{1}H-^{15}N HSQC crosspeaks move in a more symmetrical and coherent way for this protein. Possibly, this can be explained by the different binding modes of the two

proteins and that it is difficult to directly compare the K_Ds between the two proteins and binding modes.

Table 4.7 K_Ds determined by NMR for the interaction between HGF-N and different GAGs. Titration curves for each protein residue were fitted in CcpNmr Analysis after titrating the GAGs into ^{15}N-labelled HGF-N, monitoring ^{1}H-^{15}N HSQC crosspeaks. The titration curve data was then converted into a K_D in Open Office Calc and the average K_D was calculated for the residues displaying the largest CCSDs and fitting well to the equation in CcpNmr Analysis. The "number of residues" column indicates the number of residues that were used for calculating the average K_D.

GAG	Average K_D, μM	Standard Deviation	Number of residues used
Sulfated maltose	84.6	22	20
Heparin DP4C	11.8	11	19
Heparin DP4C HGF-N-Δ6	11.6	2	20
CSA DP10	41.4	15	23
CSC DP10	38.9	11	22
DS DP10	36.5	11	22

The data corresponds well to the values determined by ITC in section 4.8 and by fluorescence [21]. However, the values are several orders of magnitude different to those determined by SPR [10-13]. This is possibly due that SPR utilises immobilised ligands and suffers from inherent difficulties of characterising protein – GAG interactions due to the multivalency of GAG chains. The K_Ds of CSA, CSC and DS DP10 are very similar. This is in agreement with the affinity chromatography carried out in section 4.5.1, where it was found that HGF-N elutes approximately at the same time from CSA, CSC and DS affinity columns. It should be however emphasised that the decasaccharides used in the experiments were only size fractionated and are

therefore a potentially are mixtures of species with different level and position of sulfates. The obtained K_D values must therefore be treated conservatively.

The K_Ds for heparin DP10 and Fondaparinux could not be determined because several crosspeaks do not move on a symmetrical straight line. This is possibly due to a secondary process, such as the protein dimerising. As the binding and dimerisation is a two-phase process, it is difficult to come up with a valid model that the titration can be fitted to. Using a model for bivalent ligands [299] to fit titration data in the GraphPad software was unsuccessful.

In general, NMR is not the method of choice for K_D determination as only relatively weak interactions can be studied and that it is a relatively labour and cost intensive method [300]. On the other hand, coupled to mapping the binding interface and studying the dynamics of the interaction, it is a powerful technique that yields a lot of data in one major experiment. Compared to other techniques, NMR has the advantage of being a solution based technique and it typically requires less material than ITC, making it a competitive method for the right kind of interaction.

4.10.5 Possible Dimerisation With Heparin DP10

When ^{1}H-^{15}N HSQC spectra of heparin DP10 titrated into ^{15}N labelled HGF-N were studied, it soon became apparent that several crosspeaks do not move on a straight line between titration points, some move in an U-shape, some in a L-shape and some move on a straight line and then turn around and move back in the opposite direction. This behaviour can also be seen for Fondaparinux, but to a much smaller extent. The fact that the crosspeaks do not move on a straight line implies that the chemical environment of the backbone amides does not change in the same manner throughout the titration. This implies that two or more different processes are taking place during the titration. This behaviour cannot be seen for heparin DP4C or any other titrated ligands and it is significantly less prominent for HGF-N-Δ6 titrated with heparin DP10.

Generally, the crosspeaks for heparin DP10 move on a relatively straight line up to approximately 0.5:1 GAG:protein ratio and change direction above this ratio. This behaviour could be due to that the protein is dimerising in the beginning of the titration when there is excess protein over GAG. After this point, the probability of dimerisation decreases and monomeric species become more prevalent. Residue number 72 is a good example of a turning residue. Its crosspeak also becomes stronger during the titration, suggesting that the protein dynamics change upon GAG binding (Figure 4.93).

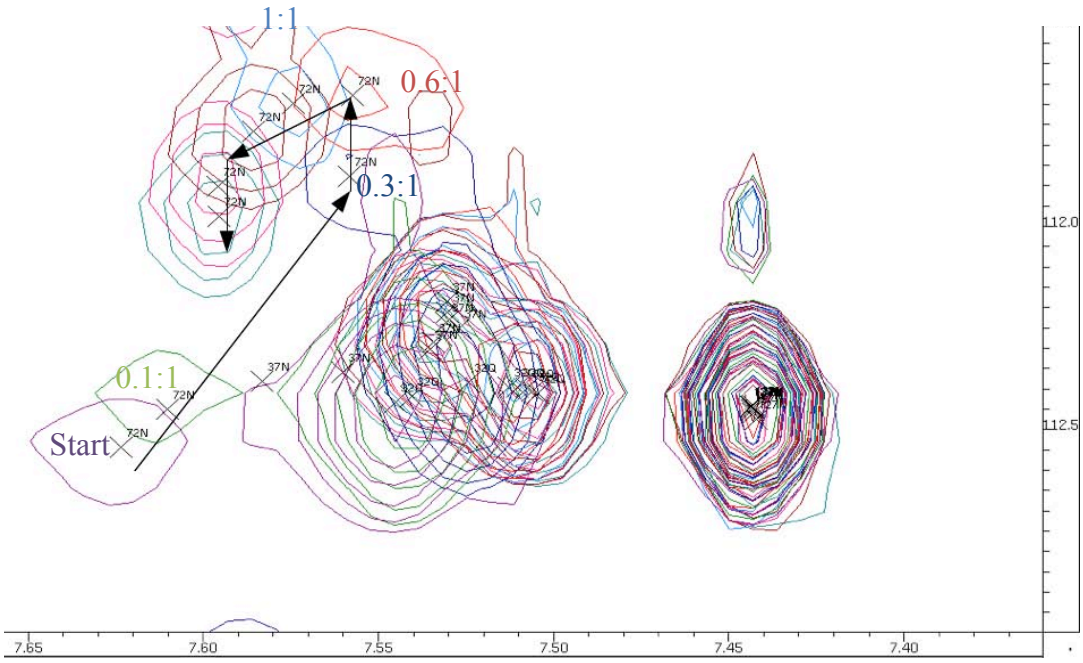


Figure 4.93 Overlaid ^{1}H-^{15}N HSQC spectra showing how the 72N crosspeak does not move on a straight line during the course of a heparin DP10 titration and how the crosspeak becomes stronger.

The less frequent occurrence of turning peaks for HGF-N-Δ6 suggests that the N-terminal residues play a role in the potential dimerisation process. Figure 4.94 shows an expansion of a ^{1}H-^{15}N HSQC spectrum showing the different behaviours of crosspeaks in the titration of DP10 into HGF-N and HGF-N-Δ6.

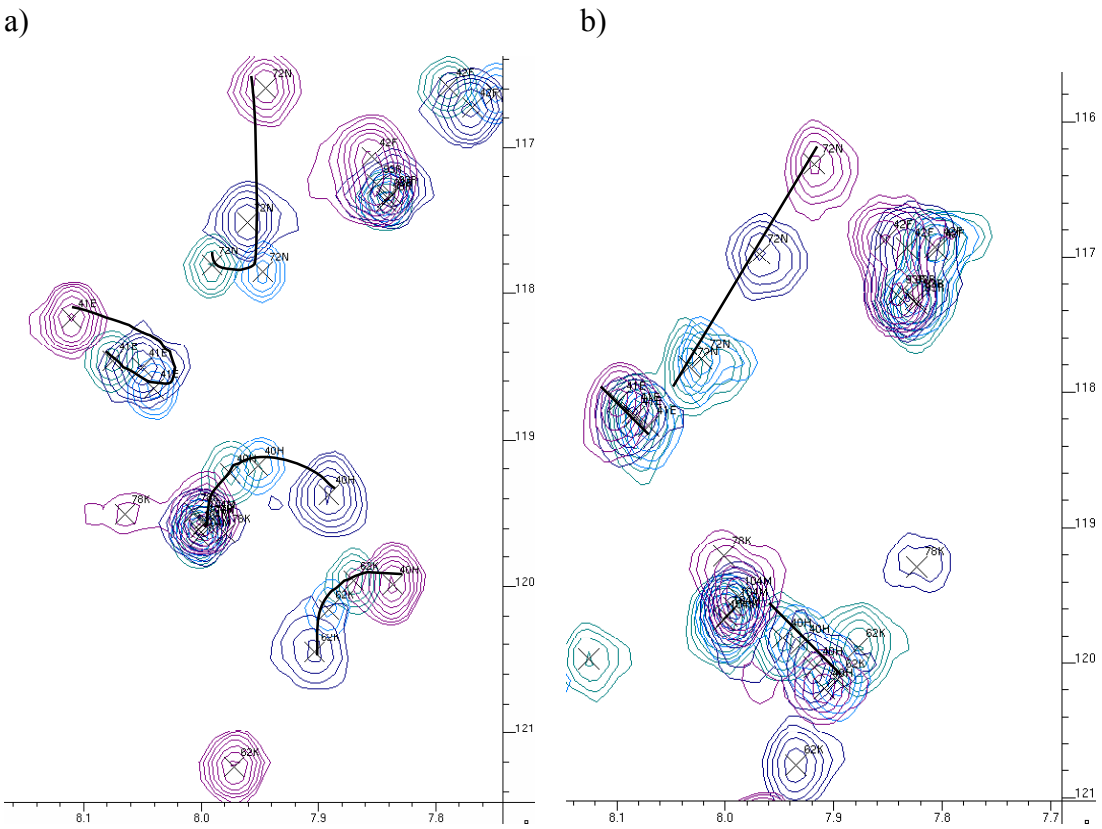


Figure 4.94 ^{1}H-^{15}N HSQC spectra illustrating how residues 40, 41, 63 and 72 move on a curved line for native HGF-N (a), but not for HGF-N-Δ6 (b) as the proteins are titrated with heparin DP10. GAG:protein ratios: a) purple: Starting point, blue: 0.3:1, light blue: 1:1, green: 10:1. b) purple: Starting point, blue: 0.2:1, light blue: 0.8:1, green: 6.4:1.

The crosspeaks that do not move on straight lines belong to residues 37-42 and 68-76 (70 and 74 moving straight for heparin DP10, but not for Fondaparinux). These crosspeaks correspond relatively well to the main α2-helix and N-terminal binding regions (Figure 4.95). It is notable that the L2 loop residues (77-84), neighbouring the turning peaks of the main α2-helix binding region, are moving on a straight line, as are the residues in the β2/4-strands binding region (59-64, 66, 95, 96).

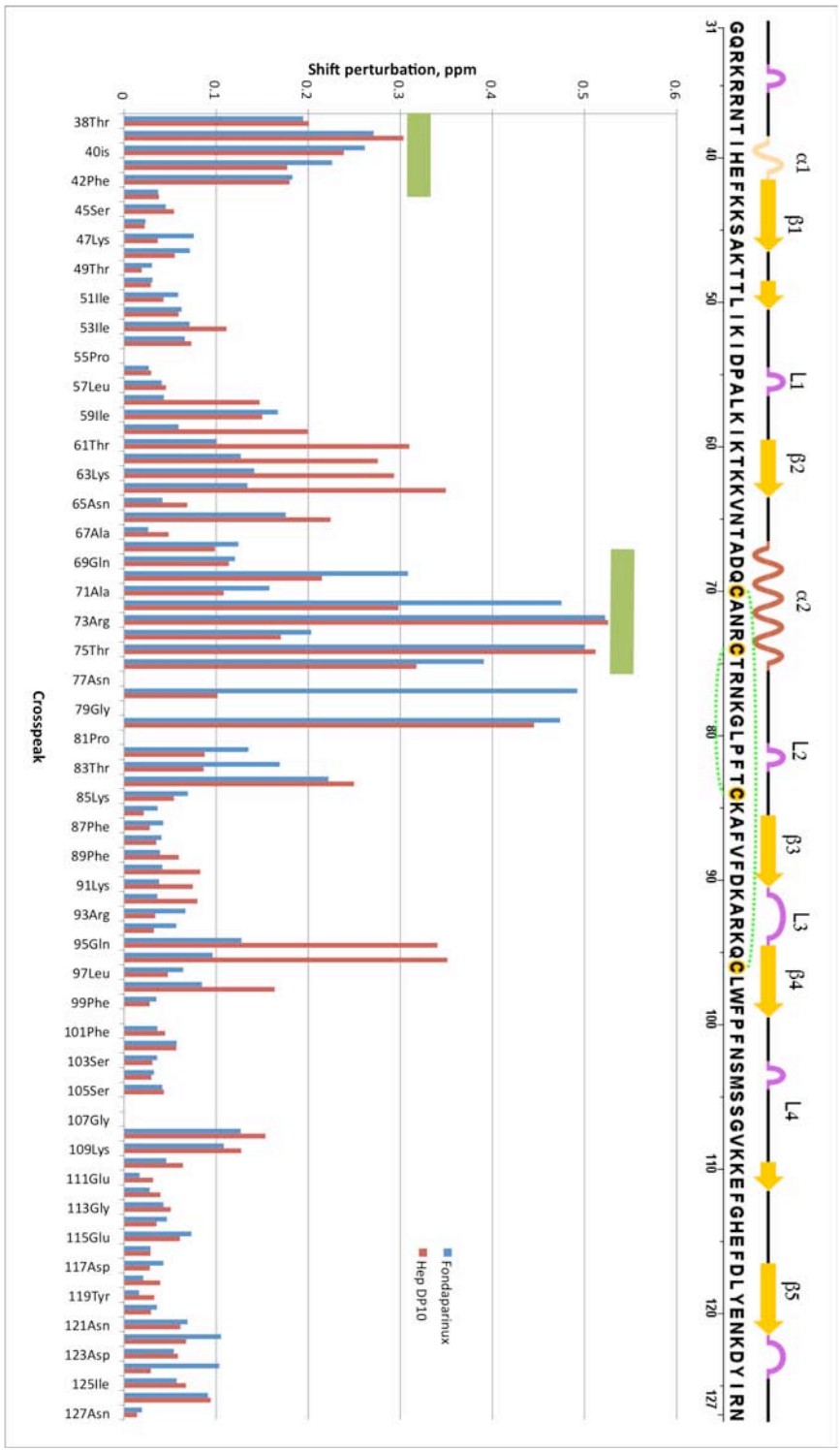


Figure 4.95 Histogram showing Fondaparinux and heparin decasaccharide titrated into HGF-N monitored by ^{1}H-^{15}N HSQC spectra. The crosspeaks that do not move on a straight line are indicated with a green line. Combined chemical shift perturbation is displayed on the y-axis and residue number on the x-axis.

The turning crosspeaks belong to residues that form a consistent surface-exposed patch, spanning the main α2-helix and N-terminal binding regions (Figure 4.96). This implies that this patch is most affected by the potential dimerisation and could be in the dimerisation interface.

a) b)

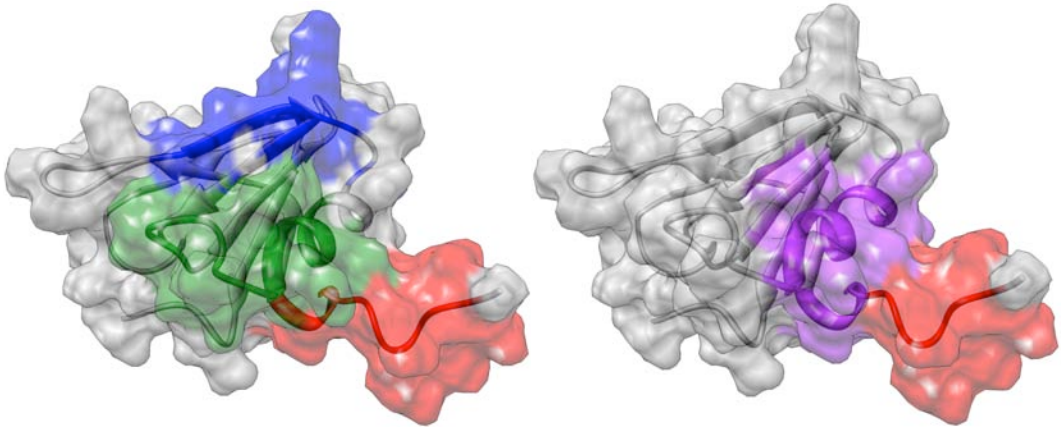


Figure 4.96 Comparison of the binding site regions and the turning peaks for heparin DP10 mapped onto the 2HGF PDB file using UCSF Chimera. a) The same as Figure 4.91; the N-terminal binding region (32-42) is coloured red, the main α2-helix binding region (70-84) is coloured green and the β2/4-strands binding region (59-64, 66, 95, and 96) is in coloured in blue. b) The N-terminal residues that cannot be seen by NMR (32-37) are coloured red and the crosspeaks that change direction during the titration (residues 38-42 and 68-76) are coloured purple.

It is likely that heparin DP10 is able to host two HGF-N molecules and makes the protein dimerise in a *cis* or *trans* mode, binding along the GAG chain (Figure 4.97). Judging from what residues that are affected by binding and the possible dimerisation, it is likely that there is an overlap between the protein-protein interface and the protein-GAG interface (Figure 4.95 and Figure 4.96).

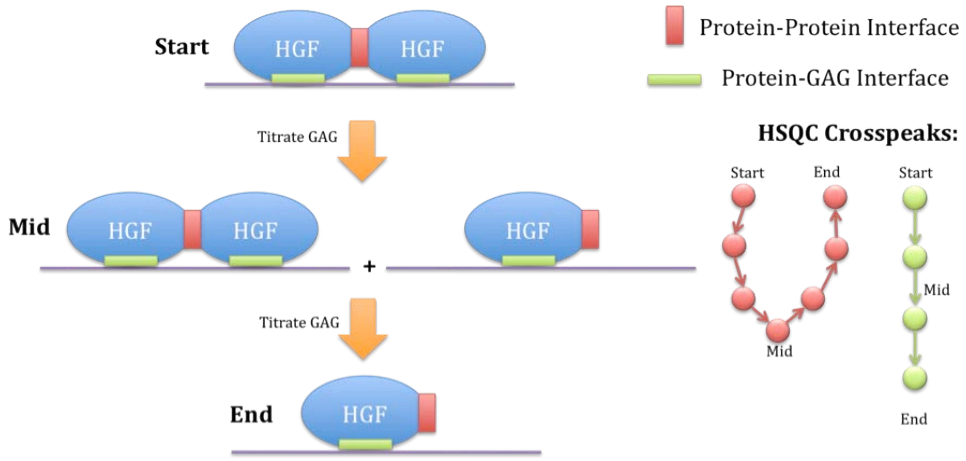


Figure 4.97 Explanation of turning ^{1}H-^{15}N HSQC crosspeaks during NMR titrations of heparin DP10.

The only evidence for that oligomerisation takes place at the beginning of titrations of HGF-N with Fondaparinux comes from NMR experiments. No clear evidence for dimerisation with Fondaparinux as ligand was seen by AUC or ITC. This could be due to the fact that NMR is the only technique able to monitor binding on an amino acid level. Although the same crosspeaks did not move on a straight line for both Fondaparinux and heparin DP10 (with the exception of residues 70 and 74), this behaviour was less obvious for Fondaparinux than for heparin DP10 suggesting that if a dimer is formed, it is less abundant in the case of Fondaparinux.

It is possible that the turning crosspeaks are not due to dimerisation, but a different process. This could for example be the existence of two or more binding modes with different affinities and the observed changes could be due to the existence of more complex equilibria than would correspond to a simple binary process. Although the results are difficult to interpret, they are reported on and discussed. Further studies will be required to fully explain these observations.

4.11 Titration of TEMPO Conjugated Heparin Hexasaccharide into ^{15}N Labelled HGF-N Monitored by NMR

GAGs display reducing end/nonreducing end directionality, just like other biological polymers such as DNA and proteins display 5'/3' and N-terminus/C-terminus directionality, respectively. The interacting points (for example sulfate groups and arginines) of a protein-GAG interaction will have different positions depending on in what direction the GAG is presented to the protein. It is therefore likely that many protein-GAG interactions display directional specificity. This directional specificity is difficult to study with many biophysical methods, including normal NMR techniques such as ^{1}H-^{15}N HSQC monitored titrations. One way of addressing this question is by attaching a spin-label to the reducing end of the GAG. In these experiments, a heparin oligosaccharide was spin-labelled and titrated into HGF-N to study the directionality of the binding event. The aim of the experiments was to produce a spin-labelled heparin oligosaccharide purified to homogeneity and titrate it into ^{15}N-labelled HGF-N, monitoring NMR relaxation times to study the directionality of the complex. The results are important in that they are the first of its kind for HGF, confirming the directionality of the HGF-NK-G – heparin co-crystal structure.

A TEMPO spin label was attached to the reducing end of a fully sulfated heparin-derived hexasaccharide by reductive amination. The spin labelled hexasaccharide was titrated into ^{15}N labelled HGF-N and chemical shift perturbation and NH proton spin-lattice relaxation rates ($R_1 = 1/T_1$) values were monitored. A reduction in the relaxation times of NH protons indicates spatial proximity of the free electron radical of specific residues. This information was used to determine the orientation in which the oligosaccharide binds to HGF-N. After the residues with enhanced relaxation were mapped onto the structure of HGF-N it became clear that the heparin-derived hexasaccharide bound preferentially in one direction.

Similar approaches have been employed to label saccharides with TEMPO and study their interaction with proteins. TEMPO-labelled N-acetyllactosamine was used to study its binding to galectin-3 and it was found that the carbohydrate binds in a specific direction and manner [301]. Another study probed the interaction between β-1,4-glucanase CenC and oligosaccharides and found that the oligosaccharide can bind in both directions due to structural symmetry of the oligosaccharide [302].

Although GAGs and carbohydrates have a direction, just like DNA, it is difficult to know whether the GAG binds in a specific direction to a protein. For example, molecular docking of protein - heparin interactions using AutoDock "cannot distinguish between 180° pairs of heparin orientations" [303]. This is likely due to the crude nature of the AutoDock method and the relative structural symmetry of the sulfate pattern along fully sulfated heparin regardless the direction.

4.11.1 Reductive Amination

There are many ways to conjugate different functionalities and reporter groups to carbohydrates [268], [304]. Reductive amination is one way of attaching an amine or hydrazide-containing molecule to the reducing end of a carbohydrate. It has the advantage of being relatively straightforward and the disadvantages of opening the reducing end monosaccharide and producing a relatively flexible linker.

In reductive amination, a free amine reacts with a carbonyl group to form an imine [268] (Figure 4.98). The imine can then be reduced with a suitable reducing agent such as sodium cyanoborohydride to form an amine. In the case of glycans, the reducing end contains a masked carbonyl in the form of a ketal or acetal group, depending on if the sugar is a ketose or aldose. The carbonyl only exists in the open ring form, which slows down the reaction.

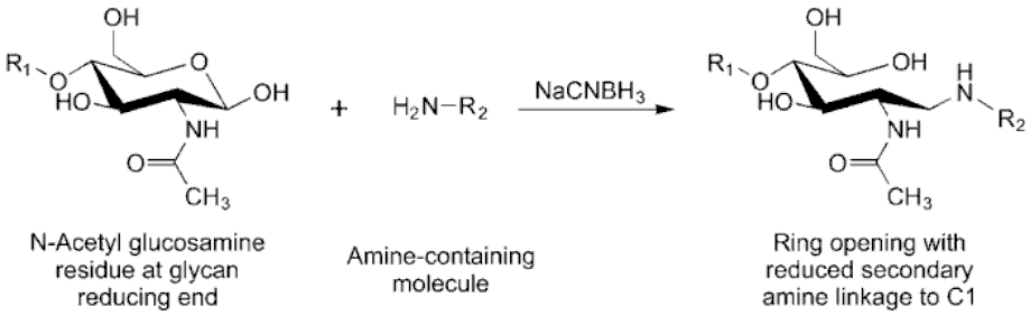


Figure 4.98 Reductive amination of a glycan. The amine attacks the carbonyl and an imine intermediate is reduced with sodium cyanoborohydride to form a conjugate. Taken from [268].

Hydrazides are more reactive and can be employed in the same reaction instead of an amine to speed it up (Figure 4.99). If the product is not reduced, an unstable hydrazide compound is formed that can hydrolyse in the presence of water [305].

Reductive amination has been widely employed to conjugate glycans to affinity labels [251], chromophores [168], etc.

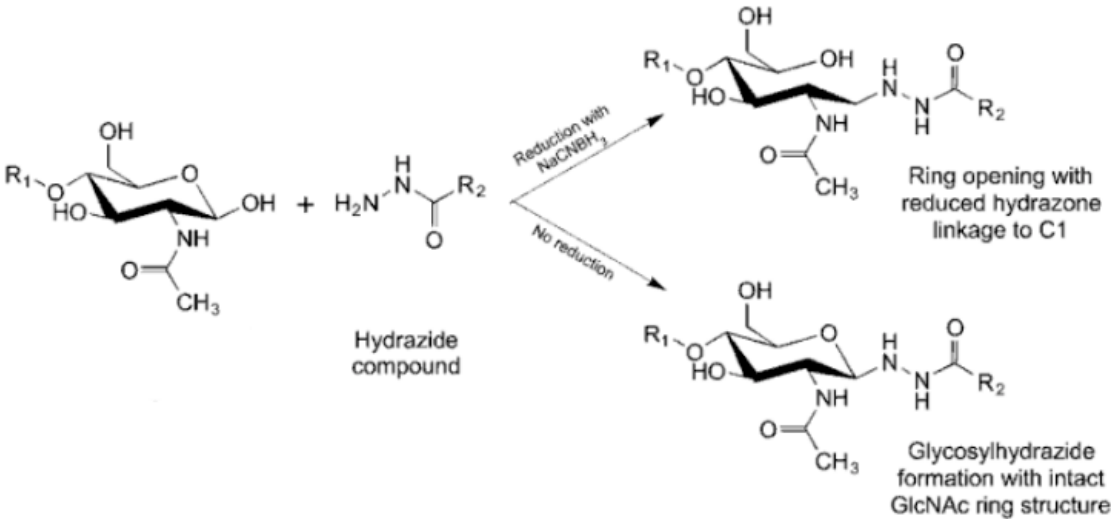


Figure 4.99 Reductive amination of a glycan with a hydrazide. Taken from [268].

A different reaction that also can be used for similar purposes is the Kochetkov amination [304], which replaces the C1 hydroxyl with an amine group that then can be used for conjugation. The reaction has the advantage of not opening the reducing end ring, preserving the ring conformation and rigidity, but the disadvantage of being very inefficient and taking several days to complete. Kochetkov amination has been employed to conjugate TEMPO to N-acetyllactosamine to characterise its interaction with galectin-3 by NMR [301].

4.11.2 Production, Purification and Analysis of the TEMPO Conjugate

To produce large quantities of heparin-derived fully sulfated hexasaccharide, an automated semi-preparative SAX chromatography method was set up on the BioCAD FPLC system. The SAX purified fully sulfated heparin DP6 was desalted, freeze-dried and conjugated to TEMPO as described in section 3.12.

After the conjugation reaction, the reaction mixture was desalted and purified by SAX chromatography, see Figure 4.100.

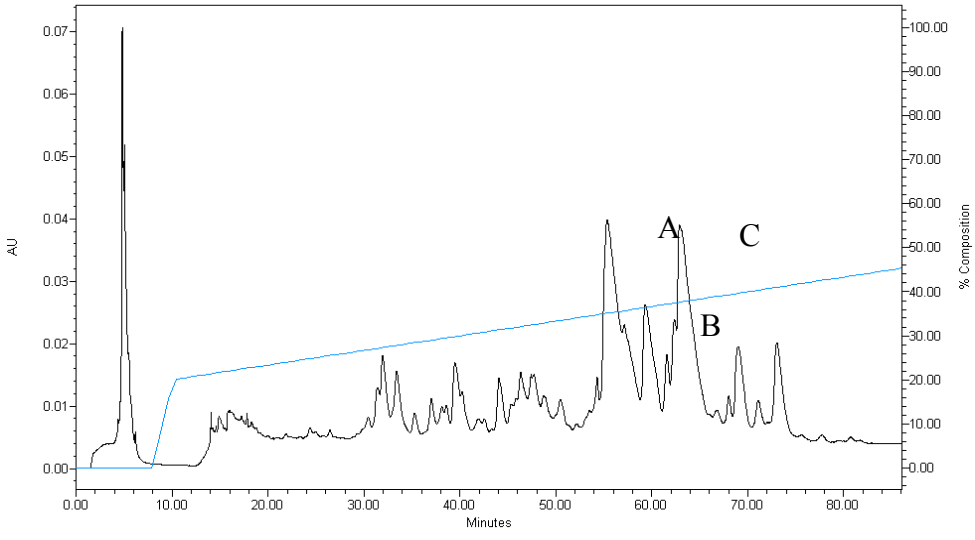


Figure 4.100 SAX chromatogram after the reductive amination of fully sulfated heparin-derived hexasaccharide by 4-Amino TEMPO. The labelled peaks contain the following species: A: DP6-TEMPO, B: Mixture of A and C, C: DP6.

Fractions from SAX chromatography were desalted and analysed by MALDI-TOF mass spectrometry (Figure 4.101 and Figure 4.102). Chromatography peaks eluting between 30-50 minutes did not contain any GAGs. The peak eluting at 55 minutes contained fully sulfated heparin DP6 with a conjugated TEMPO. The peak eluting at 60 minutes contained a mixture of unconjugated and conjugated DP6. The peak eluting at 63 minutes contained unconjugated DP6. Peaks eluting after 63 minutes did not contain any GAGs. The mass shift between the two species in Figure 4.102 is 155.7 Da. The mass of 4-Amino TEMPO is 171.3 Da, a water molecule is lost in imine formation and two hydrogens are donated by the reducing agent, which means that the final mass error is 0.4 Da, which is acceptable as no internal standards were used and the spectrum was not deisotoped. The peak eluting at 55 minutes was desalted, freeze-dried and used for the protein titration.

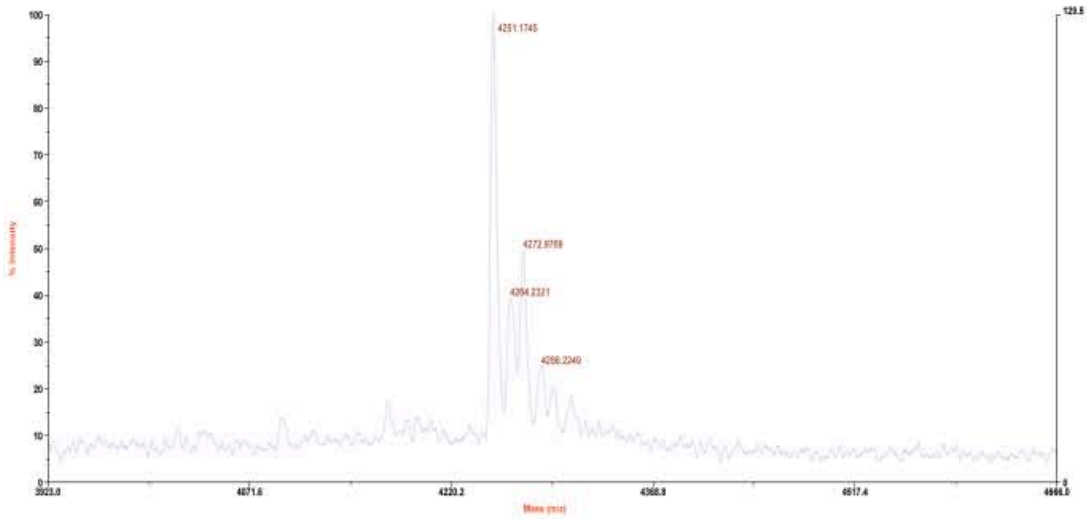


Figure 4.101 MALDI-TOF spectrum of the pure TEMPO-DP6 species eluting from the SAX column at 55 minutes (peak A).

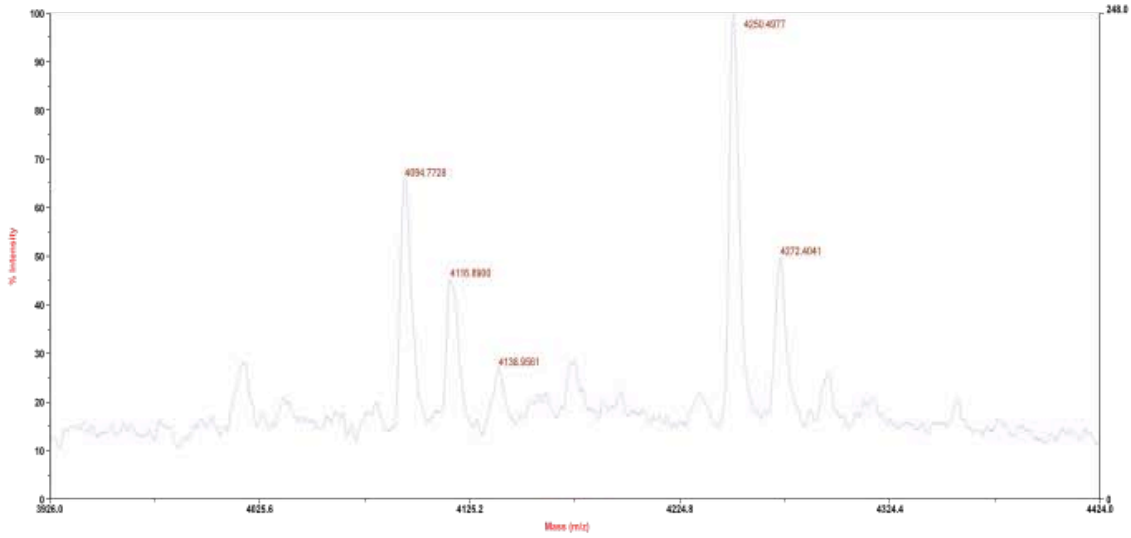


Figure 4.102 MALDI-TOF spectrum of the SAX peak eluting at 60 minutes (peak B) containing both conjugated and unconjugated DP6 with typical sodium adducts.

To further assess purity and structural integrity of the conjugate, its ^{1}H NMR spectrum was collected and compared with the starting material (Figure 4.103).

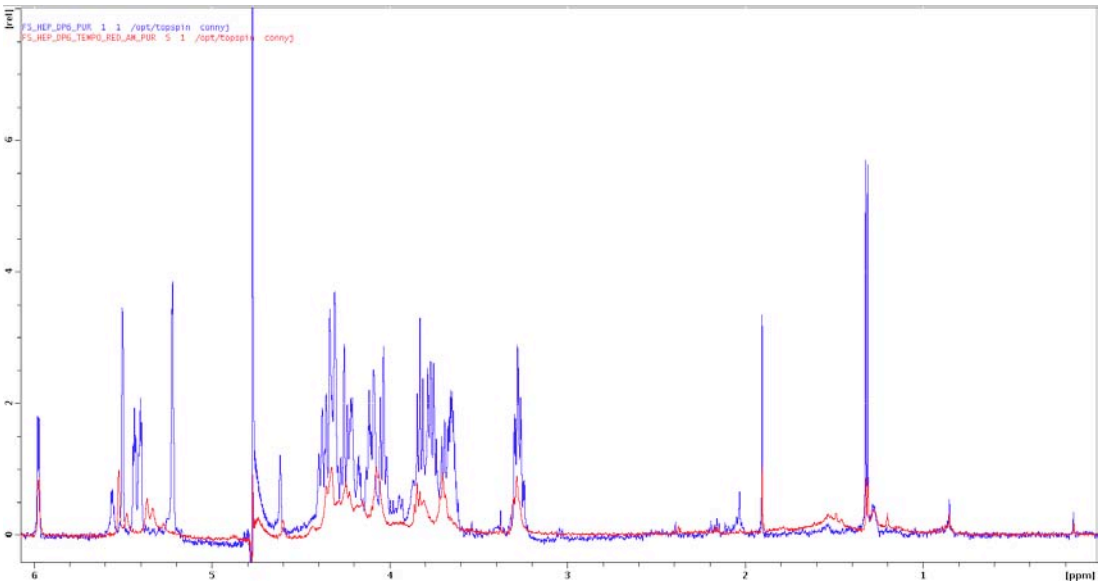


Figure 4.103 Overlaid ^{1}H NMR spectra showing fully sulfated heparin DP6 in blue and its TEMPO conjugate in red.

New signals can be seen in the spectrum of the conjugate at around 1.5 ppm. This is the region where TEMPO signals resonate. However, the radical broadens the signals and they should not be visible at all if they were from the TEMPO spin label. Possibly, the new signals are from a contaminant. Nevertheless, carbohydrate signals dominate the spectrum, indicating sufficient purity of the sample and its stability for further experiments.

To further confirm that the spin label was attached to DP6, ^{1}H T_1 inversion recovery spectra were collected (see Table 4.8 for relaxation delay times). Spectra were collected for both the starting material and the conjugate (Figure 4.104 and Figure 4.105, respectively).

Table 4.8 Inversion recovery times used for the calculation of ^{1}H T_1 relaxation spectra.

Spectrum No	**Relaxation Delay (s)**	**Estimated T_1 based on zero signal**
1	0.15	0.21
2	0.2	0.28
3	0.25	0.35
4	0.3	0.42
5	0.4	0.56
6	0.5	0.7
7	0.6	0.84
8	0.8	1.12
9	1	1.4

It can be seen that most of the signals in the DP6 have close to zero intensity in spectra 4, 5, and 6. Based on this approximation, the proton relaxation times of heparin DP6 can be estimated to be in the range of 0.4 to 0.7 s. This estimate is based on the relationship between relaxation delay and T_1, where $T_1 = \sqrt{2}RD$, where RD is the relaxation delay that produces zero intensity signals.

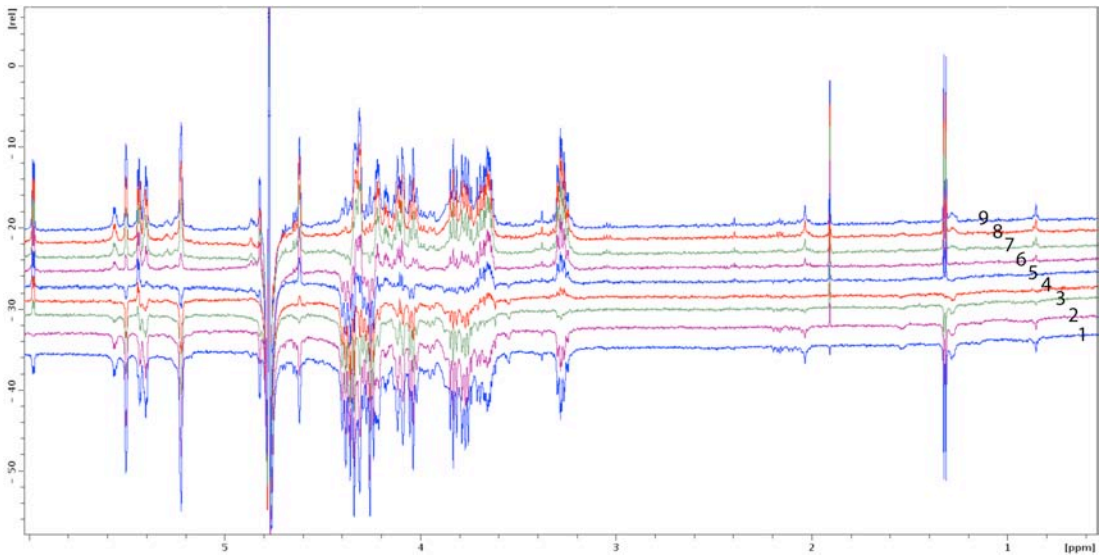


Figure 4.104 ^{1}H T_1 relaxation spectra for starting material, fully sulfated heparin DP6, see Table 4.8 for delay times.

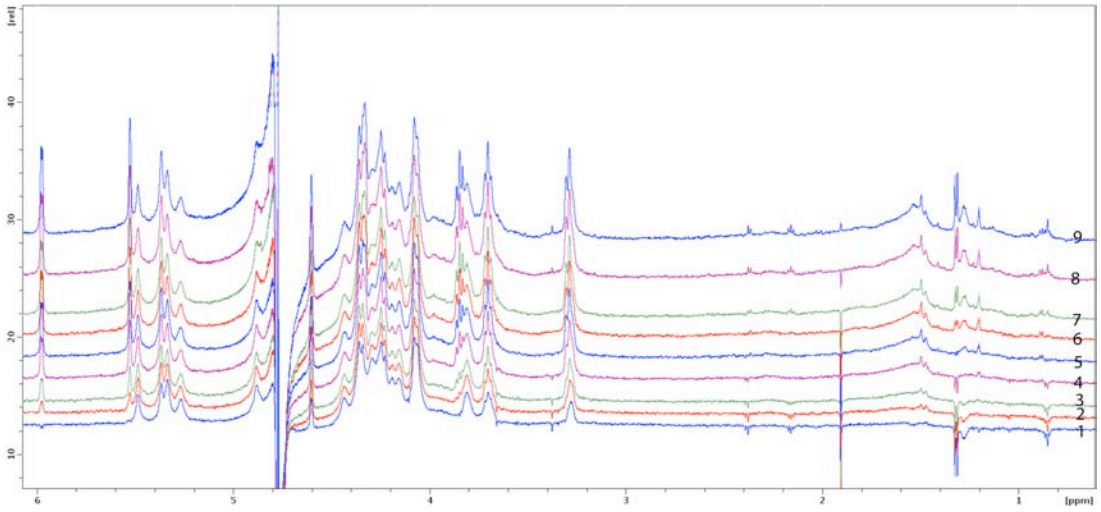


Figure 4.105 ^{1}H T_1 relaxation spectra for TEMPO-DP6, see Table 4.8 for delay times.

As expected, the protons of the conjugate relaxed much faster than those of the starting material. Note that all GAG signals are positive already in the first spectrum of the TEMPO-DP6 conjugate shown in Figure 4.105. This is due to the spin label enhancing the relaxation of the protons. Because all the signals are positive, the

relaxation times cannot be determined, but it can be said that the T_1 is below 0.21 s. The signal at 6 ppm in Figure 4.105 and Figure 4.106 belongs to the double bond proton of the non-reducing end monosaccharide of the hexasaccharide. Comparing this signal in both sets of spectra, it is clear that this proton is less affected by the spin-label than the other protons. Its T_1 relaxation time changed only from 0.28 s to 0.21 s. This is because the paramagnetic relaxation enhancement decreases quickly with the increasing distance from the spin label according to equation 4.1 [301] and that this proton is furthest away from the spin label.

$$\Delta\left(\frac{1}{T}\right) = \frac{2K}{r^6}\left(\frac{3\tau_c}{1+\omega_H^2\tau_c^2}\right) \qquad \textbf{(4.1)}$$

where τ_c is the correlation time for the molecule, ω_H is the Larmor frequency for the proton in question, T is the longitudinal relaxation time, r is the distance of the nucleus from the unpaired electron, and K is a collection of fundamental constants related to spin properties of the system.

The doublet signal at 1.3 ppm seen in both DP6 and TEMPO-DP6 spectra is likely a contaminant and it is reassuring to see that it is not as affected by the TEMPO spin label as the GAG signals. For this signal, its estimated T_1 relaxation time actually increased from 0.42 s to 0.56 s. We can therefore argue that the TEMPO-DP6 did not cause non-specific relaxation at the concentration used.

In addition to the faster T_1 relaxation, T_2 values of the protons of TEMPO-DP6 are also smaller than those of the pure DP6. This is manifested in the broader lines of the ^1H spectrum of the conjugate.

It was concluded that the TEMPO was conjugated to DP6 and that it was possible to purify the conjugate by SAX chromatography. NMR and MS analysis showed the expected masses and NMR spectral features of the molecule in Figure 4.106.

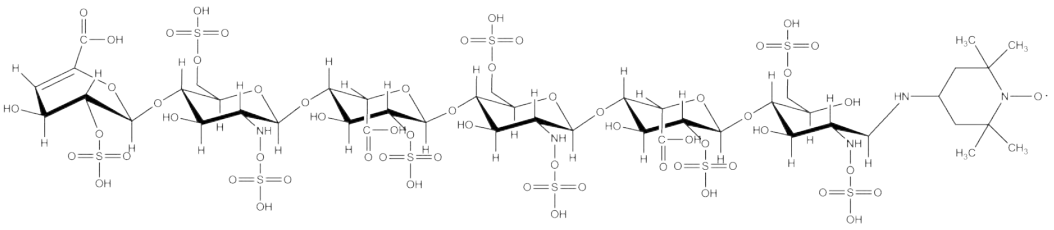


Figure 4.106 Fully sulfated heparin hexasaccharide, conjugated with 4-amino TEMPO by reductive amination. The molecule is fully protonated as in acidic mass spectrometry conditions.

4.11.3 NMR Titration of HGF-N and TEMPO Conjugated Fully Sulfated Heparin-Derived Hexasaccharide

As the TEMPO-DP6 was titrated into HGF-N, both chemical shift changes and intensity changes of the crosspeaks could be observed in ^{1}H-^{15}N HSQC spectra (Figure 4.107). The chemical shift differences are due to the binding of the TEMPO-DP6 and reflect changes in the chemical environment of the backbone NH protons. The intensity changes are caused by the enhanced relaxation of the backbone NH protons due to the presence of TEMPO or due to chemical exchange. Intensity changes due to chemical exchange are generally accompanied by chemical shift changes.

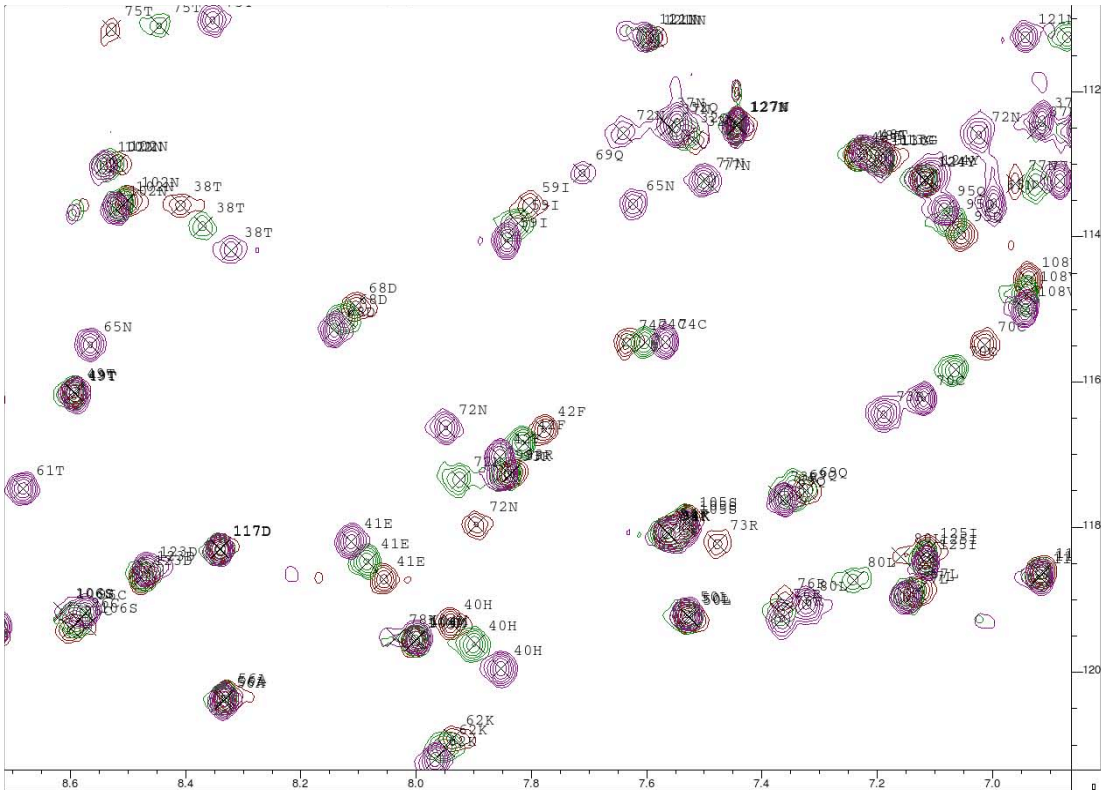


Figure 4.107 ^{1}H-^{15}N HSQC spectra of HGF-N monitored as TEMPO-DP6 is titrated. The purple spectrum is the starting point, the green is at 0.5 times excess of TEMPO-DP6 and the brown spectrum is at equimolar ratio. Note how for example 61T, 65N, 69Q and 77N quench in the presence of spin labelled GAG, while the chemical shifts of other crosspeaks such as 38T, 40H, 41E, 70C, 72N, 73R, and 80L change.

Specific binding of the TEMPO-DP6 and relaxation enhancement by the spin label rather than by chemical exchange is illustrated by comparing a residue that is not affected by the binding of the TEMPO-DP6 spin label with another that is strongly affected by the TEMPO-DP6 spin label (Figure 4.108).

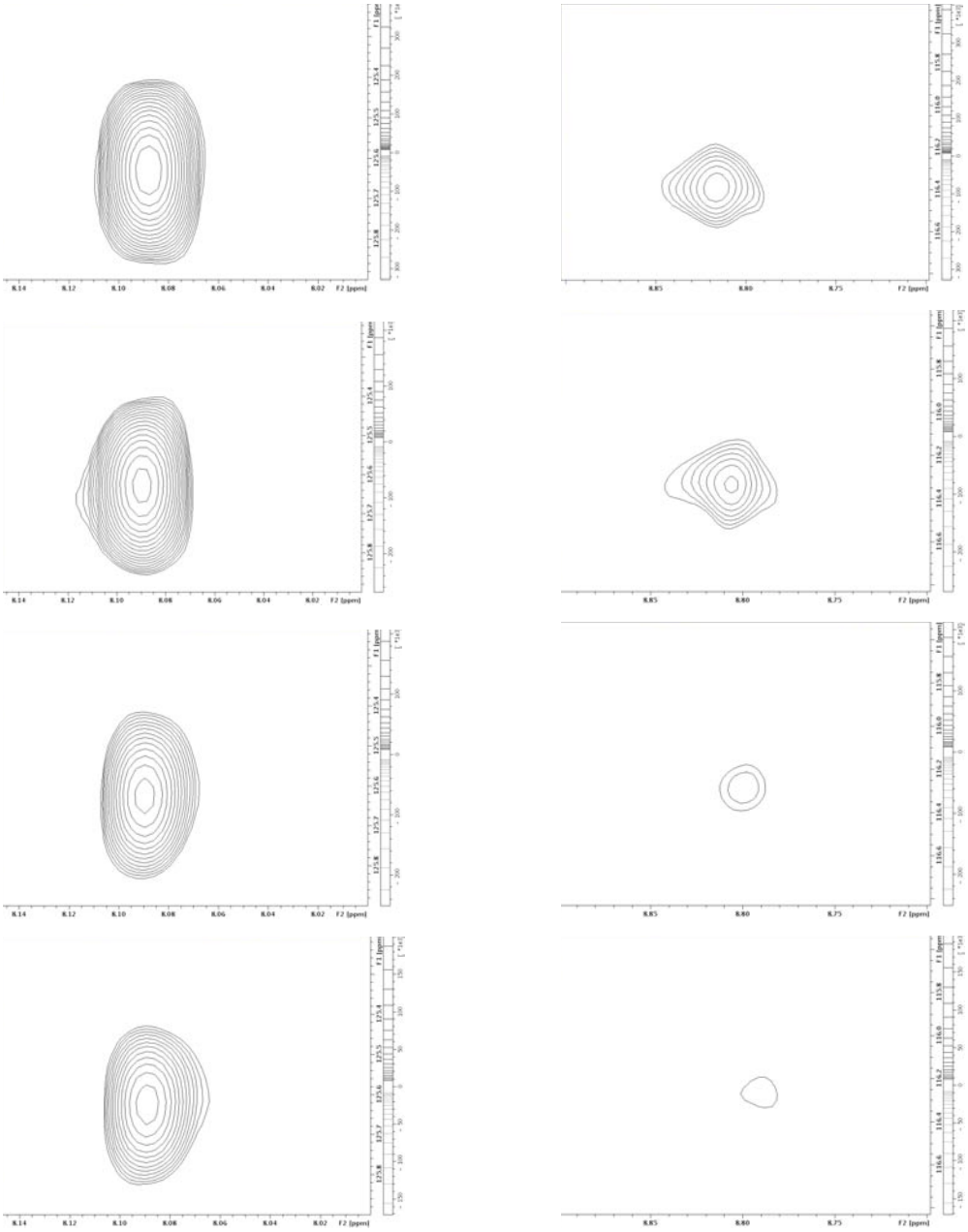


Figure 4.108 Partial ^{1}H-^{15}N HSQC spectra that show how Asn127 (on the left) is not affected by the spin label, whereas Thr83 (on the right) is affected by the spin label. TEMPO-DP6 : HGF-N ratio from the top to the bottom: 0:1, 0.5:1, 1:1 and 1.8:1.

Combined chemical shift differences of ^{1}H and ^{15}N atoms and paramagnetic contributions to the relaxation rates (ΔR_1) of NH protons were calculated in the presence of an equimolar amount of TEMPO-DP6. T_1 values of NH protons were determined from a series of relaxation experiments as described in section 3.13 for the free protein and three TEMPO-DP6 titration points. The contribution of the spin label towards the spin-lattice relaxation rates was calculated as

$$\Delta R_1 = R_1(Complex) - R_1(Free)$$

where $R_1 = \frac{1}{T_1}$, $R_1(Complex)$ = R_1 of the HGF-N – TEMPO-DP6 complex and $R_1(Free)$= R_1 of free HGF-N protein.

The combined chemical shift perturbation was expressed as $\sqrt{(\Delta\delta H)^2 + \left(\frac{\Delta\delta N}{5}\right)^2}$ as described in section 3.11.

Chemical shift perturbations and R1 values were determined in CcpNmr Analysis 2.1.2 and the data for was plotted in Open Office Calc 3.1.1 (Figure 4.109). Studying Figure 4.109, it can be noted that the ΔR1 values are elevated above zero. There are three potential reasons for this:

A) The relaxation rates increase upon binding of the hexasaccharide. This is a diamagnetic contribution not caused by the spin label.

B) The relaxation rates increase due to non-specific relaxation caused by TEMPO-DP6.

C) Free contaminating TEMPO, although this is less likely after two gel filtration and one SAX purification steps.

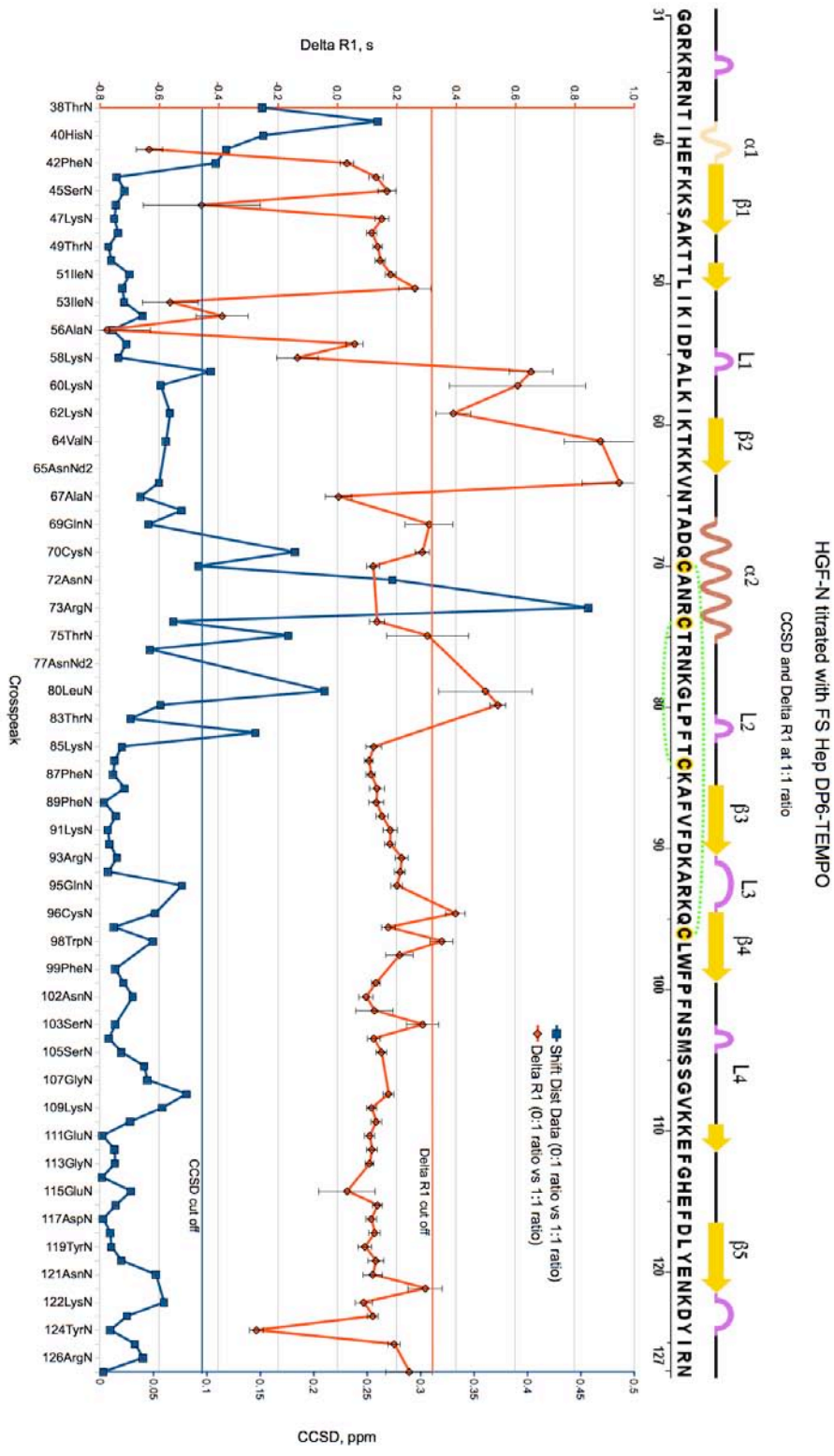


Figure 4.109 Chemical shift perturbation and ΔR_1 data. The secondary structure of HGF-N and annotation is shown on the top. The standard deviation is shown as error bars.

There are also some ΔR_1 outliers that have a negative value and some residues could not be fitted to the equation at all because of unknown reasons. The negative ΔR_1 values imply that the relaxation rates are slower in the presence of TEMPO-DP6, meaning that the residues become more flexible upon TEMPO-DP6 binding. The general school of thought is that proteins become stabilised upon ligand binding. It is likely that most parts of HGF-N are rigidified upon GAG binding, but at the same time it has been shown that the protein undergoes conformational exchange upon heparin binding and that the protein contains several flexible loops [21]. Interestingly, residues 53 to 58 all have negative ΔR_1 values and belong to a flexible loop not far away from the heparin binding site. These residues belong to the L1 loop, which is one of the most flexible parts of the protein together with the L2 loop [21]. It is therefore possible that this loop becomes even more flexible upon heparin binding to accommodate the conformational changes in the other parts of the protein. This question could be answered by determining ΔR_1 values for free HGF-N and HGF-N complexed with heparin DP6 without a spin label, but this experiment was not carried out due to time constraints.

A conservative ΔR_1 cut off value of 0.35 was chosen to select residues that had elevated relaxation rates in the presence of TEMPO-DP6. These were mapped onto protomer C of the 1GMO pdb file with UCSF Chimera 1.4 (Figure 4.111) it is clear that L1, L2 and β2/4 are more affected by the spin label than the rest of HGF-N and that the N-terminal residues and α2 helix are more affected by TEMPO-DP6 binding as evidenced by the CCSDs. This implies that TEMPO-DP6 is binding preferentially in one direction, with the reducing end monosaccharide towards the protein β2/4 sheet and L1 loop. Using the glycosidic bonds as a handle, one could also say that that TEMPO-DP6 binds in a C1 to C4 direction from the protein N-terminus.

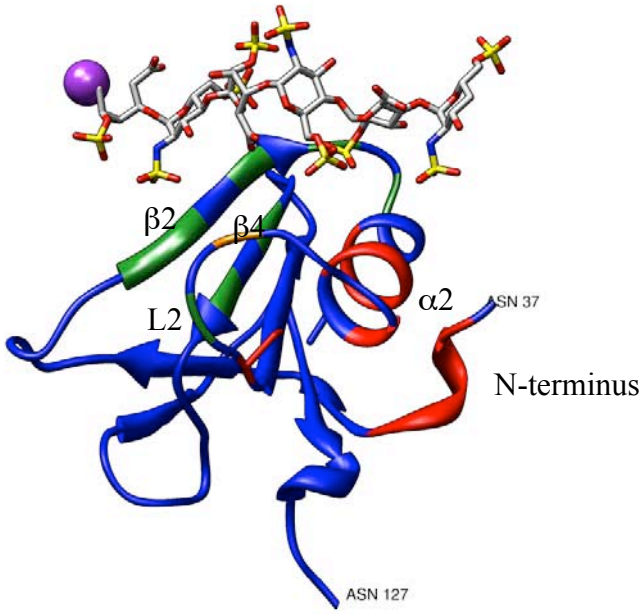


Figure 4.111 Ribbon presentation of the N domain of protomer C of the 1GMN pdb file. Residues coloured red have a shift perturbation ≥0.09 ppm, i.e. they are strongly affected by GAG binding, residues coloured green have a ΔR_1 ≥0.35 s, i.e. they are strongly affected by the TEMPO spin label. The orange residue is affected by both binding and the spin label. The TEMPO spin label is represented as a purple circle. Some areas of the green and red surfaces do not form consistent patches as ΔR_1 could not be determined for some of the residues.

To compare the results with the literature, the HGF-NK-G - heparin crystal structure was analysed [14]. This crystal structure contains several protein-heparin interfaces. It was found that in all protomers except one, the heparin oligosaccharide bound in the same direction as determined in this project. Therefore, it can be said that there is a preference for the heparin oligosaccharide to bind in a C1 to C4 direction rather than in a C4 to C1 direction (Table 4.9). This shows that although the HGF-GAG interaction is promiscuous, there is some specificity and preference for the GAG to bind in a certain direction. One can imagine that although the cell surface GAGs form a complex network *in vivo*, HGF needs to bind GAGs in a specific direction to be presented to the Met receptor, form a ternary complex and activate the receptor.

Table 4.9 GAG binding directions from the HGF-N N-terminus. The "H – B" annotation refers to that the protein-heparin interaction is taking place between the two different protomers.

Source	**Direction**
1GMO Protomer A	C4 to C1
1GMO Protomer B	C1 to C4
1GMO Protomer C	C1 to C4
1GMO Protomer F	C1 to C4
1GMO Protomer G	C1 to C4
1GMO Protomer H - B	C1 to C4
1GMN	C1 to C4
NMR experiment:	C1 to C4

One potential drawback of using spin labelled oligosaccharides to study binding events as in this experiment is that the oligosaccharide is modified on the reducing end and that this could alter the binding mode of the oligosaccharide due to steric hindrance or other effects. The spin labelled GAG will situate itself on the binding site in the most favourable way. For example, it was found in section 4.10 that the main α2-helix and N-terminal binding regions of HGF-N were of high importance for the interaction with heparin and that the β2/4-strands region was of less importance. The TEMPO-DP6 used in this study is therefore more likely to situate itself with the TEMPO group pointing towards the β2/4-strands region to maximise normal interactions between heparin and the more important main α2-helix and N-terminal binding regions.

This approach used in this experiment is applicable to many other protein-carbohydrate interactions and can help elucidating questions about binding directions

and specificities. Using a reductive amination reaction to link the spin label to a carbohydrate opens the reducing end sugar unit and produces a flexible linker to the spin label. This is not optimal for paramagnetic relaxation enhancement experiments as structural information about the reducing end sugar unit is lost and the flexible linker makes it difficult to use the spin label to measure distances. Nevertheless this experiment proves that it is possible to derive structural information using reductive amination to conjugate a spin label. Kochetkov amination and other alternative reactions can be employed to produce more rigid linkers.

It would be interesting to determine T_1 relaxation times as a normal heparin oligosaccharide is being titrated. This could potentially help to clarify the elevated relaxation times across the protein and why some residues, particularly in the L1 loop deviate from the overall ΔR_1 increase. It could also potentially clarify the conformational change that HGF-N undergoes when binding a GAG as indicated by the appearance of new ^{1}H-^{15}N HSQC peaks [21].

4.12 Future Studies

4.12.1 The role of GAGs in the Activation of HGFA and HGF

HGF and HGFA are potential therapeutic targets and the signalling cascade is currently being targeted for drug development [306-310]. Activation of HGFA by thrombin and kallikrein 4 are dependent upon the presence of a negatively charged molecule such as dextran sulfate [56], [57]. Further, it has been shown that HGFA acquires heparin affinity after it is activated by thrombin [62]. These two facts indicate that HGFA interacts with GAGs. As described in the introduction, HGFA is activated by cleavage, creating a short chain consisting of the first 25 N-terminal amino acids and has a theoretical pI of 11.71. The long chain consists of 248 amino acids with a theoretical pI of 7.22. The high pI of the short chain implies that the GAG binding site is located there. It is possible that the GAG binding site is covered in the unactivated state and exposed in the activated state after thrombin cleavage. In vivo, the unactivated HGFA is likely to be present and diffusing in many tissues, but

after tissue injury and HGFA activation it acquires heparin affinity, becomes less mobile at the injury site as it binds to GAGs, which ensures localised HGF activation. HGF's inherent GAG affinity brings it closer to the cell surface where it can be activated by HGFA and then bind the Met receptor. These examples point to the central role of GAGs in the HGF signalling cascade as shown in Figure 1.8. It would therefore interesting to investigate the activation and GAG binding of HGFA in more detail to understand the central role these central processes in Met activation and how it could be involved in different physiological functions and disease states.

There are several relevant experiments to characterise the interactions between HGFA, HGF and GAGs, for example:

A) Crystal trials with complexes of different versions and mutations of HGF and HGFA.

B) Kinetics (K_m and K_{cat}) for HGF activation by HGFA. This could for example be done by ITC and SPR.

C) NMR studies on activated and unactivated HGFA. There is no NMR assignment or dynamics studies of HGFA, which would supplement the crystal structure data and clarify the structural differences between the two different forms. Further, HGFA has been shown to undergo an extreme conformational change upon binding of its substrate and HGFA inhibitors. NMR would be an ideal technique to characterise the conformational dynamics.

D) Activation of ^{15}N-labelled HGFA by unlabelled thrombin monitored by NMR. This would clarify the structural differences between the two different forms and how heparin affinity is acquired.

E) HGFA binding to GAGs.

    a. Comparison of different GAGs.

    b. Minimal GAG oligosaccharide length.

    c. ITC, NMR and SPR binding parameters.

d. NMR structural differences in the presence and absence of GAGs.

F) The role of GAGs in HGFA activation by thrombin and kallikrein 4.

4.12.2 The Hepatocyte Growth Factor – Glycosaminoglycan Interaction

The two most well characterised protein - GAG interactions are probably the AT-III-heparin and FGF-heparin interactions. AT-III exhibits a relatively high degree of specificity, whereas FGF exhibits less specificity and oligomerises upon heparin and receptor binding. The HGF-GAG interaction is also relatively well characterised, but it seems to differ from FGF and AT-III in that it is more promiscuous and has a different oligomerisation mode, although the latter is not completely clarified yet. As the HGFA-HGF-Met signalling cascade has been implicated in many important diseases, it is relevant to investigate the HGF-GAG interaction further, not only for its involvement in disease, but also as a model for protein-GAG interactions.

As the heparin-HGF-N/NK interaction was successfully characterised by ITC, it would be interesting to extend the studies to estimate the contribution of non-ionic interactions, deduce how many sodium ions that are released and how many protons that are transferred in the binding event [282]. This would help clarifying how many and what kind of intermolecular bonds that are formed in the binding event. It would also be interesting to further compare the difference between different oligosaccharide lengths. With the GAG oligosaccharide preparation method that was setup in this project and the good expression levels of HGF-N, sufficient material for ITC studies can easily be achieved. The MicroCal Auto VP-ITC instrument is well suited for these kinds of experiments as it is equipped with an auto-sampler.

It was shown in this project that the N-terminal residues of the HGF N domain are important for GAG binding. Further, it was shown that it is likely that the HGF-NK-G protein construct used in the HGF-NK-G-heparin co-crystal is binding in a different mode to heparin due to an N-terminal cloning artefact. To clarify the

differences between the native HGF N domain and the HGF-NK-G protein construct, it would be interesting to solve the crystal or NMR structure of a protein construct with a native N-terminus in the presence of a heparin oligosaccharide. Solving the NMR structure in the presence of heparin/Fondaparinux could also help clarifying the occurrence of new ^{1}H-^{15}N HSQC in titrations of these ligands.

Although HGF-N has been relatively well characterised by NMR, T_1/T_2/NOE data in the absence of a GAG oligosaccharide, these parameter have still not been established in the presence of a GAG oligosaccharide ligand. The investigation of these relaxation parameters would help clarifying the conformational change that the protein undergoes upon binding of a GAG as has been shown previously [21] and in this project. For example, data in section 4.11.3 suggested that the L2 loop becomes more flexible upon ligand binding.

The results from the gel filtration studies of HGF-N and HGF-NK-G complexed with CS, DS and heparin oligosaccharides of varying lengths showed that all GAG oligosaccharides bound and it was pointed out that the large hydrodynamic radius of GAGs has to be taken into consideration when interpreting the data. A more quantitative approach that involves calculating hydrodynamic properties of the different oligomerisation modes seen in Figure 4.59, using for example the HYDROPRO software [233], and comparing them with the experimental values would give deeper insight into the interaction.

5 Conclusions

Protein – GAG interactions are involved in the regulation of a wide range of physiological processes such as blood coagulation, tissue regeneration and embryogenesis. Hepatocyte growth factor regulates physiological processes as diverse as cell growth, survival, proliferation, chemotaxis, cell morphology, tissue regeneration and angiogenesis. In this project, the interactions between different GAGs and several protein constructs of the GAG binding domains of HGF have been characterised using a range of biophysical techniques.

The three different protein constructs (HGF-N, HGF-K and HGF-NK-C) were cloned into an expression vector for *P. pastoris* from human cDNA, expressed in a bench-top fermenter and purified to homogeneity. The protein construct HGF-NK-G was obtained from a collaborator. Expression levels of HGF-N and HGF-K were relatively good with a 5-litre fermentation yielding more than 100 mg of one-step purified protein. Some of the expressed HGF-N protein was shown by mass spectrometry to have N-terminal truncations due to faulty signal-peptide cleavage during secretion. Expression of HGF-NK-C proved difficult due to N-terminal truncations and two protease sites, which were mapped by mass spectrometry. Inclusion of peptone or tryptone in the expression medium allowed unlabelled HGF-NK-C to be expressed in useable amounts. It can be concluded that protein production was successful and most hurdles were overcome. *P. pastoris* was an effective protein expression system for the investigated protein constructs, but N-terminal truncations and proteolysis decreased the yields.

To compare the binding affinity between the different protein constructs and GAGs, CSA, CSC, DS and heparin hexadecasaccharides were biotinylated and immobilised on Sepharose-streptavidin columns, effectively producing GAG affinity columns (section 4.5.1). The results showed that all investigated protein constructs bind with the highest affinity to a standard GE Healthcare HiTrap heparin column, followed by the in-house produced heparin column and with the lowest but similar affinities to

CSA, CSC and DS -columns. The affinity of HGF-N for hyaluronan was judged to be non-significant under physiological conditions. To my knowledge, it has previously not been shown that HGF binds to CS. In these experiments, HGF-N, HGF-NK-C and HGF-NK-G bound with similar affinity to CS and DS and the interaction is thought to be physiologically relevant as the proteins elute from the GAG affinity column at approximately 200-400 mM NaCl. The results align well with the gel filtration results in section 4.6, where it was found that the investigated protein constructs bind to both CS and DS under gel filtration conditions in PBS. Different preparations of DS was used in section 4.6 and 4.5.1, which likely explains why the general observation was that the affinity was higher for DS in the gel filtration experiments and higher for CS in the affinity chromatography experiments.

ITC was employed to study the thermodynamics of the interactions between HGF-N and HGF-NK-G protein constructs and Fondaparinux and different heparin oligosaccharide lengths. The ITC results show that the entropy contribution was favourable for short oligosaccharides and disfavourable for long oligosaccharides and that the enthalpy contribution was smaller for short oligosaccharides than for long oligosaccharides. The opposite trends in enthalpy and entropy as a function of oligosaccharide length explain why the dissociation constants do not change much with varied oligosaccharide length. It has previously been shown by SPR that the affinity of HGF for heparin-derived oligosaccharides does not change with length [10]. To my knowledge, this is the first time ITC has been used to study the HGF-GAG interaction and the entropy and enthalpy contribution to the binding has been determined. The results help explain the affinity of the HGF-GAG interaction and why long oligosaccharides do not necessarily bind tighter than short oligosaccharides. This knowledge can aid in drug design of GAG mimetics. The experiments can be extended by performing ITC titrations in buffers with different ionic strengths and heats of ionisation. By using different buffers, the contribution of non-ionic interactions can be determined as well as how many sodium ions that are released in the binding event and how many protons are transferred.

The ITC data showed that the dissociation constant between HGF-N, HGF-NK-G, different heparin oligosaccharide lengths and Fondaparinux varied between 0.35 μM and 9.26 μM. Dissociation constants determined by NMR and ITC for the interaction in this project were in the same range as previously reported for the HGF-N - heparin interaction by solution-based fluorescence methods [21]. However, the dissociation constants between different HGF constructs and GAGs determined by SPR are in a low nM range in many previous studies [10-13]. Thus the dissociation constants determined by NMR and ITC in this project deviate several orders of magnitude from those determined by SPR in other studies. The main reason for the large difference in dissociation constants is likely that NMR and ITC are solution based techniques, whereas SPR requires ligand immobilisation on a surface, which can give rise to artefacts. In addition, it is often difficult to correct for the multivalency of protein-GAG interactions in SPR experiments. SPR has been the most common method to study the HGF-GAG interaction and surprisingly few alternative solution-based methods have been used. The results presented in this thesis, in particular those obtained by the ITC and NMR experiments, are therefore important. They indicate that solution-based methods could be more appropriate for studies of the HGF-GAG interaction. More solution-based techniques need to be employed in studying the HGF-GAG interaction in order to clarify the wide range of K_Ds determined by different methods.

NMR titrations of CSA, CSC, DS, heparin -decasaccharides, heparin tetrasaccharide, Fondaparinux and chemically sulfated maltose into ^{15}N labelled HGF-N allowed the GAG binding site to be mapped and studied in detail. It is shown that all ligands bind to the same general HGF-N binding site, but different binding modes exists between the different ligands. Three binding regions were identified within the binding site, with the α2-helix and L2 loops playing key roles (residues 70-84). All GAGs also utilise a N-terminal binding region (residues 32-42), whereas long heparin oligosaccharides can also utilise a binding region formed mainly by the β2 and β4 - strands (residues 59-64, 66, 95, 96). The latter binding region is more important when the first six N-terminal residues are removed (HGF-N-Δ6) compared to the

native HGF-N protein. It is likely that this binding mode is also present in HGF-NK-G, which has a non-native N-terminus. The importance of these residues was also confirmed by heparin affinity chromatography: both HGF-N-Δ6 and HGF-NK-G eluted from heparin affinity columns earlier than HGF-N and HGF-NK-C. The results are in line with previously published data [63] and are important as they provide detailed insight into the importance of the N-terminal binding region of HGF, which can't be studied by NMR or X-ray methods due to its flexibility. To my knowledge, this is also the first time a repertoire of GAGs has been used to compare different binding modes to HGF on a molecular level. The detailed mapping of HGF's GAG binding site in conjunction with the ITC results in this thesis improves the understanding of this important interaction and can aid in the development of GAG mimetics.

Thermal denaturation of HGF-N monitored by NMR showed that heparin binding stabilises HGF-N; the melting temperature of HGF-N increases in the presence of heparin decasaccharide. These novel results improve the basic understanding of the HGF-N construct as its melting temperature is determined and it is shown that HGF-N is stabilised by heparin decasaccharide.

A number of protein–carbohydrate interactions exhibit directional specificity [301], whereas others proteins can bind its ligand in any direction [302]. To investigate whether HGF-N exhibits directional specificity, a spin-label was attached to the reducing end of fully sulfated heparin hexasaccharide. The effect of the spin label on HGF-N was examined by determining the CCSDs and relaxation times of the protein backbone NH protons by NMR. After the data was mapped onto the structure of HGF-N it became clear that the heparin-derived hexasaccharide bound preferably in one direction. Although the HGF-GAG interaction appears to be relatively promiscuous as the protein can bind several different types of GAGs and accommodate a variety of sulfation patterns, the results suggest that the protein preferably binds heparin and HS in one direction. This is not surprising because even though HGF-N has the ability to bind different GAGs and sulfation patterns, the

protein is likely to exhibit some specificity and preference for sugar backbone sequences and sulfation patterns. In the same manner, the protein is likely to exhibit directional preference as the positions of sulfates and other interaction points will clearly change depending on what direction the GAG binds. The results are novel in that it is the first time that binding directionality is studied for the HGF-GAG interaction by NMR. The results show the same orientation of the GAG as seen in the X-ray structure of the HGF-NK-G - heparin complex. The X-ray structure is based on the protein with a cloning artefact, which affects the binding as demonstrated in our work. It was therefore important to investigate whether the direction of the binding is the same for the HGF-N protein construct and in solution. The directionality of GAG binding is important for the design of GAG mimetics, when targeting the HGF-GAG interaction for drug development. The results are also important in that it is shown that reductive amination can be used to attach spin-labels to GAGs, derive structural information and determine the directionality of their binding to proteins. Reductive amination gives rise to an opened sugar unit and a flexible linker on the reducing end where the spin-label is attached. This is not optimal for paramagnetic relaxation enhancement experiments as structural information is lost in the ring opening and the flexible linker makes it more difficult to determine the position of the spin label accurately.

It has previously been suggested that HGF binds preferably highly sulfated GAG species containing iduronic acid. To investigate what heparin sequence in a hexasaccharide mixture that binds tightest to HGF-N, an affinity column was prepared by biotinylating the protein construct and immobilising it on a GE Healthcare Streptavidin HiTrap column. A mixture of enzymatically derived heparin hexasaccharides was loaded onto the column and eluted with a salt gradient. The affinity-purified hexasaccharides were characterised by SAX chromatography and MALDI-TOF mass spectrometry. The results were in line with previously published studies [67]: HGF-N bound tightest to fully sulfated heparin hexasaccharide that was likely built up by IdoA sugar units. Somewhat unexpectedly, no 3-O sulfated hexasaccharide was detected, likely due to the low abundance of 3-O sulfate in

heparin and relative insensitivity of the method. Hexasaccharides lacking more than three sulfates did not bind under the experimental conditions. It appears that HGF, just like many other heparin binding proteins, is able to accommodate a wide range of sulfation patterns and degrees of sulfation, although a preference for highly sulfated species exists. These results are important as they show that fluorescent tag labelling of GAGs is not required for their direct detection after affinity chromatography.

A self-associating dimer of HGF-N that bound with extraordinary high affinity to heparin was characterised by heparin affinity chromatography, gel filtration and NMR. Heparin affinity chromatography showed that dimeric HGF-N binds tighter to heparin than monomeric HGF-N. Gel filtration proved that the tight binding species was indeed a dimer that is relatively stable at 4 °C, but dissociates into a monomer if incubated at 37 °C. The dissociation of the dimeric species was studied by NMR at 30 °C and as it took several hours for the dimer to dissociate, it can be concluded that the off-rate of the complex must be very slow. What induces dimerisation is unknown. The dissociation of the complex is slow on an NMR timescale, which makes it difficult to follow the NH crosspeaks in H-^{15}N HSQC spectra. Some crosspeaks that could be assigned with satisfactory confidence were mapped on the protein structure. The mapped surface does not correspond to any protein-protein interfaces seen in the crystal structures of HGF. Further studies have to be conducted to clarify the nature of this complex, for example, the binding interface can be mapped by mass spectrometry methods. HGF-NK-G crystallises as a head-to-tail dimer, but the protein is likely monomeric in solution. These results are novel as a HGF-N dimer has not been reported before. Further studies could clarify if the higher affinity is solely due to the dimer having two GAG binding sites or if a more complex mechanism is involved.

Gel filtration is a common technique used to study the oligomeric state of protein-GAG complexes. However, due to the very large hydrodynamic radii of GAGs, it is difficult to interpret experiments when a mixture of a protein and GAG is analysed

by gel filtration. Based on previous gel filtration experiments and crystal structures, it has been suggested that HGF-NK oligomerises upon the addition of GAGs. To clarify the contribution of the GAG to the elution time in gel filtration studies, free GAGs and free proteins were analysed separately by gel filtration before analysing the mixtures. The results showed that free CSC DP12 migrated as an apparent larger molecular weight than free HGF-N and that free CSC DP22 migrated as an apparent larger molecular weight than free HGF-NK-G. This emphasises the importance of taking the hydrodynamic radius of the GAG into consideration when performing gel filtration experiments of protein-GAG complexes. It also stresses how difficult it is to interpret the results as a change in protein oligomeric state will partly be hidden by the large hydrodynamic radius of the GAG. Nevertheless, when taking the GAG radius into consideration, evidence for a 2:1 protein:GAG complex could be observed for heparin chains that were long enough to accommodate two protein molecules (≥DP10 for HGF-N and ≥DP14 for HGF-NK-G), implying that HGF-N and HGF-NK-G have a simple oligomerisation mode like beads on a string. AUC analysis was employed to confirm the gel filtration results and it was found that free HGF-N and HGF-NK-G are monomeric in solution and in the presence of equimolar Fondaparinux. In the presence of equimolar heparin decahexasaccharide, higher-order species could be detected, likely representing two proteins bound to one decahexasaccharide. The AUC results also suggested a simple beads on a string oligomerisation mode. ITC analysis further confirmed the beads on a string model, showing 1:1 interactions for Fondaparinux and heparin hexasaccharide, but a 2:1 protein:GAG interaction for heparin decahexasaccharide. The results are important in that they provide good evidence for that the HGF-N and HGF-NK protein construct have a simple beads on a string oligomerisation mode, which is a novel finding that increases our basic understanding of how these protein constructs interact with GAGs.

6 References

[1] W. Jiang, "Hepatocyte growth factor and the hepatocyte growth factor receptor signalling complex as targets in cancer therapies," *Current Oncology*, vol. 14, no. 2, Apr. 2007.

[2] L. E. Kemp, B. Mulloy, and E. Gherardi, "Signalling by HGF/SF and Met: the role of heparan sulphate co-receptors.," *Biochemical Society Transactions*, vol. 34, no. 3, pp. 414-417, Jun. 2006.

[3] A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, G. W. Hart, and M. E. Etzler, *Essentials of Glycobiology, Second Edition*, 2nd ed. Cold Spring Harbor Laboratory Press, 2008.

[4] B. Casu, A. Naggi, and G. Torri, "Heparin-derived heparan sulfate mimics that modulate inflammation and cancer," *Matrix Biology*.

[5] V. Ferro et al., "PI-88 and novel heparan sulfate mimetics inhibit angiogenesis," *Seminars in Thrombosis and Hemostasis*, vol. 33, no. 5, pp. 557-68, Jul. 2007.

[6] J. K. Fairweather, E. Hammond, K. D. Johnstone, and V. Ferro, "Synthesis and heparanase inhibitory activity of sulfated mannooligosaccharides related to the antiangiogenic agent PI-88," *Bioorganic & Medicinal Chemistry*, vol. 16, no. 2, pp. 699-709, Jan. 2008.

[7] C. R. Parish, C. Freeman, K. J. Brown, D. J. Francis, and W. B. Cowden, "Identification of Sulfated Oligosaccharide-based Inhibitors of Tumor Growth and Metastasis Using Novel in Vitro Assays for Angiogenesis and Heparanase Activity," *Cancer Research*, vol. 59, no. 14, pp. 3433-3441, Jul. 1999.

[8] L. M. Khachigian and C. R. Parish, "Phosphomannopentaose sulfate (PI-88): heparan sulfate mimetic with clinical potential in multiple vascular pathologies," *Cardiovascular Drug Reviews*, vol. 22, no. 1, pp. 1-6, 2004.

[9] T. A. Martin and W. G. Jiang, "Hepatocyte growth factor and its receptor signalling complex as targets in cancer therapy," *Anti-Cancer Agents in Medicinal Chemistry*, vol. 10, no. 1, pp. 2-6, Jan. 2010.

[10] M. Delehedde, M. Lyon, R. Vidyasagar, T. J. McDonnell, and D. G. Fernig, "Hepatocyte Growth Factor/Scatter Factor Binds to Small Heparin-derived Oligosaccharides and Stimulates the Proliferation of Human HaCaT Keratinocytes," *Journal of Biological Chemistry*, vol. 277, no. 14, pp. 12456-12462, Mar. 2002.

[11] G. Hartmann et al., "Engineered mutants of HGF/SF with reduced binding to heparan sulphate proteoglycans, decreased clearance and enhanced activity in vivo," *Current Biology*, vol. 8, no. 3, pp. 125-135, Jan. 1998.

[12] M. Lyon, J. A. Deakin, H. Rahmoune, D. G. Fernig, T. Nakamura, and J. T. Gallagher, "Hepatocyte Growth Factor/Scatter Factor Binds with High Affinity to Dermatan Sulfate," *Journal of Biological Chemistry*, vol. 273, no. 1, pp. 271-278, Jan. 1998.

[13] T. Merkulova-Rainon, P. England, S. Ding, C. Demerens, and G. Tobelem, "The N-terminal Domain of Hepatocyte Growth Factor Inhibits the Angiogenic Behavior of Endothelial Cells Independently from Binding to the c-met Receptor," *Journal of Biological Chemistry*, vol. 278, no. 39, pp. 37400-37408, Sep. 2003.

[14] D. Lietha, D. Y. Chirgadze, B. Mulloy, T. L. Blundell, and E. Gherardi, "Crystal structures of NK1–heparin complexes reveal the basis for NK1 activity and enable engineering of potent agonists of the MET receptor," *The EMBO Journal*, vol. 20, no. 20, pp. 5543–5555, Oct. 2001.

[15] M. Ultsch, N. A. Lokker, P. J. Godowski, and A. M. de Vos, "Crystal structure of the NK1 fragment of human hepatocyte growth factor at 2.0 A resolution," *Structure*, vol. 6, no. 11, pp. 1383-1393, Nov. 1998.

[16] D. Y. Chirgadze, J. P. Hepple, H. Zhou, R. A. Byrd, T. L. Blundell, and E. Gherardi, "Crystal structure of the NK1 fragment of HGF/SF suggests a novel mode for growth factor dimerization and receptor binding," *Nature Structural and Molecular Biology*, vol. 6, no. 1, pp. 72-79, Jan. 1999.

[17] H. Zhou et al., "The solution structure of the N-terminal domain of hepatocyte growth factor reveals a potential heparin-binding site," *Structure*, vol. 6, no. 1,

pp. 109-116, Jan. 1998.

[18] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 4th ed. Garland Science, 2002.

[19] V. Cioce et al., "Hepatocyte Growth Factor (HGF)/NK1 Is a Naturally Occurring HGF/Scatter Factor Variant with Partial Agonist/Antagonist Activity," *Journal of Biological Chemistry*, vol. 271, no. 22, pp. 13110-13115, May. 1996.

[20] K. Watanabe, D. Y. Chirgadze, D. Lietha, H. de Jonge, T. L. Blundell, and E. Gherardi, "A New Crystal Form of the NK1 Splice Variant of HGF/SF Demonstrates Extensive Hinge Movement and Suggests That the NK1 Dimer Originates by Domain Swapping," *Journal of Molecular Biology*, vol. 319, no. 2, pp. 283-288, May. 2002.

[21] H. Zhou et al., "Identification and Dynamics of a Heparin-Binding Site in Hepatocyte Growth Factor," *Biochemistry*, vol. 38, no. 45, pp. 14793-14802, Nov. 1999.

[22] K. Miyazawa, "Hepatocyte growth factor activator (HGFA): a serine protease that links tissue injury to activation of hepatocyte growth factor," *FEBS Journal*, vol. 277, no. 10, pp. 2208-2214, 2010.

[23] J. Stamos, R. A. Lazarus, X. Yao, D. Kirchhofer, and C. Wiesmann, "Crystal structure of the HGF [beta]-chain in complex with the Sema domain of the Met receptor," *EMBO Journal*, vol. 23, no. 12, pp. 2325-2335, Jun. 2004.

[24] D. Kirchhofer et al., "Structural and Functional Basis of the Serine Protease-like Hepatocyte Growth Factor {beta}-Chain in Met Binding and Signaling," *Journal of Biological Chemistry*, vol. 279, no. 38, pp. 39915-39924, Sep. 2004.

[25] W. D. Tolbert, J. Daugherty-Holtrop, E. Gherardi, G. Vande Woude, and H. E. Xu, "Structural basis for agonism and antagonism of hepatocyte growth factor," *Proceedings of the National Academy of Sciences*, Jul. 2010.

[26] A. Weiss and J. Schlessinger, "Switching Signals On or Off by Receptor

Dimerization," *Cell*, vol. 94, no. 3, pp. 277-280, Aug. 1998.

[27] M. Tallquist and A. Kazlauskas, "PDGF signaling in cells and mice," *Cytokine & Growth Factor Reviews*, vol. 15, no. 4, pp. 205-213, Aug. 2004.

[28] N. J. Harmer, L. L. Ilag, B. Mulloy, L. Pellegrini, C. V. Robinson, and T. L. Blundell, "Towards a Resolution of the Stoichiometry of the Fibroblast Growth Factor (FGF)-FGF Receptor-Heparin Complex," *Journal of Molecular Biology*, vol. 339, no. 4, pp. 821-834, Jun. 2004.

[29] C. J. Robinson, N. J. Harmer, S. J. Goodger, T. L. Blundell, and J. T. Gallagher, "Cooperative Dimerization of Fibroblast Growth Factor 1 (FGF1) upon a Single Heparin Saccharide May Drive the Formation of 2:2:1 FGF1·FGFR2c·Heparin Ternary Complexes," *Journal of Biological Chemistry*, vol. 280, no. 51, pp. 42274-42282, Dec. 2005.

[30] K. Saxena et al., "Influence of heparin mimetics on the assembly of the FGF - FGFR4 signaling complex," *Journal of Biological Chemistry*, 2010.

[31] A. Canales et al., "Solution Nmr Structure of a Human Fgf-1 Monomer, Activated by a Hexasaccharide Heparin-Analogue," *FEBS Journal*, vol. 273, no. 20, pp. 4716-4727, Oct. 2006.

[32] J. P. Dawson, M. B. Berger, C. Lin, J. Schlessinger, M. A. Lemmon, and K. M. Ferguson, "Epidermal Growth Factor Receptor Dimerization and Activation Require Ligand-Induced Conformational Changes in the Dimer Interface," *Molecular and Cellular Biology*, vol. 25, no. 17, Sep. 2005.

[33] M. A. Olayioye, R. M. Neve, H. A. Lane, and N. E. Hynes, "The ErbB signaling network: receptor heterodimerization in development and cancer," *EMBO Journal*, vol. 19, no. 13, pp. 3159-3167, Jul. 2000.

[34] R. Tao and I. N. Maruyama, "All EGF(ErbB) receptors have preformed homo- and heterodimeric structures in living cells," *Journal of Cellular Science*, vol. 121, no. 19, pp. 3207-3217, Oct. 2008.

[35] M. Miller and E. J. Leonard, "Mode of receptor binding and activation by plasminogen-related growth factors," *FEBS Letters*, vol. 429, no. 1, pp. 1-3,

Jun. 1998.

[36] E. Gherardi et al., “Functional map and domain structure of MET, the product of the c-met protooncogene and receptor for hepatocyte growth factor/scatter factor,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 21, pp. 12039-12044, Oct. 2003.

[37] M. Lyon, J. A. Deakin, and J. T. Gallagher, “The Mode of Action of Heparan and Dermatan Sulfates in the Regulation of Hepatocyte Growth Factor/Scatter Factor,” *Journal of Biological Chemistry*, vol. 277, no. 2, pp. 1040-1046, Jan. 2002.

[38] T. F. Zioncheck et al., “Sulfated Oligosaccharides Promote Hepatocyte Growth Factor Association and Govern Its Mitogenic Activity,” *Journal of Biological Chemistry*, vol. 270, no. 28, pp. 16871-16878, Jul. 1995.

[39] J. S. Rubin et al., “Dissociation of Heparan Sulfate and Receptor Binding Domains of Hepatocyte Growth Factor Reveals That Heparan Sulfate-c-Met Interaction Facilitates Signaling,” *Journal of Biological Chemistry*, vol. 276, pp. 32977 -32983, 2001.

[40] D. Bechard et al., “Endocan Is a Novel Chondroitin Sulfate/Dermatan Sulfate Proteoglycan That Promotes Hepatocyte Growth Factor/Scatter Factor Mitogenic Activity,” *Journal of Biological Chemistry*, vol. 276, no. 51, pp. 48341-48349, Dec. 2001.

[41] J. Villena and E. Brandan, “Dermatan sulfate exerts an enhanced growth factor response on skeletal muscle satellite cell proliferation and migration,” *Journal of Cellular Physiology*, vol. 198, no. 2, pp. 169-178, Feb. 2004.

[42] O. Holmes et al., “Insights into the Structure/Function of Hepatocyte Growth Factor/Scatter Factor from Studies with Individual Domains,” *Journal of Molecular Biology*, vol. 367, no. 2, pp. 395-408, Mar. 2007.

[43] G. Hartmann et al., “A functional domain in the heavy chain of scatter factor/hepatocyte growth factor binds the c-Met receptor and induces cell dissociation but not mitogenesis.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 23, pp. 11574-11578,

Dec. 1992.

[44] D. Kirchhofer et al., "Utilizing the activation mechanism of serine proteases to engineer hepatocyte growth factor into a Met antagonist," *Proceedings of the National Academy of Sciences*, vol. 104, no. 13, pp. 5306-5311, Mar. 2007.

[45] W. D. Tolbert et al., "A mechanistic basis for converting a receptor tyrosine kinase agonist to an antagonist," *Proceedings of the National Academy of Sciences*, p. 0704290104, Sep. 2007.

[46] S. Goldoni et al., "Decorin is a novel antagonistic ligand of the Met receptor," *J. Cell Biol.*, vol. 185, no. 4, pp. 743-754, May. 2009.

[47] M. Peek, P. Moran, N. Mendoza, D. Wickramasinghe, and D. Kirchhofer, "Unusual Proteolytic Activation of Pro-hepatocyte Growth Factor by Plasma Kallikrein and Coagulation Factor XIa," *Journal of Biological Chemistry*, vol. 277, no. 49, pp. 47804-47809, Nov. 2002.

[48] K. Miyazawa, T. Shimomura, A. Kitamura, J. Kondo, Y. Morimoto, and N. Kitamura, "Molecular cloning and sequence analysis of the cDNA for a human serine protease reponsible for activation of hepatocyte growth factor. Structural similarity of the protease precursor to blood coagulation factor XII," *Journal of Biological Chemistry*, vol. 268, no. 14, pp. 10024-10028, May. 1993.

[49] K. A. Owen et al., "Pericellular activation of hepatocyte growth factor by the transmembrane serine proteases matriptase and hepsin, but not by the membrane-associated protease uPA," *The Biochemical Journal*, vol. 426, no. 2, pp. 219-228, Mar. 2010.

[50] S. Lee, R. B. Dickson, and C. Lin, "Activation of Hepatocyte Growth Factor and Urokinase/Plasminogen Activator by Matriptase, an Epithelial Membrane Serine Protease," *Journal of Biological Chemistry*, vol. 275, no. 47, pp. 36720-36725, Nov. 2000.

[51] H. Kataoka and M. Kawaguchi, "Hepatocyte growth factor activator (HGFA): pathophysiological functions in vivo," *FEBS Journal*, vol. 277, no. 10, pp. 2230-2237, 2010.

[52] H. Kataoka, S. Miyata, S. Uchinokura, and H. Itoh, “Roles of hepatocyte growth factor (HGF) activator and HGF activator inhibitor in the pericellular activation of HGF/scatter factor,” *Cancer and Metastasis Reviews*, vol. 22, no. 2, pp. 223-236, Jun. 2003.

[53] E. Gherardi et al., “Structural basis of hepatocyte growth factor/scatter factor and MET signalling,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 11, pp. 4046-4051, Mar. 2006.

[54] S. Shia et al., “Conformational lability in serine protease active sites: structures of hepatocyte growth factor activator (HGFA) alone and with the inhibitory domain from HGFA inhibitor-1B,” *Journal of Molecular Biology*, vol. 346, no. 5, pp. 1335-49, Mar. 2005.

[55] C. Eigenbrot, R. Ganesan, and D. Kirchhofer, “Hepatocyte growth factor activator (HGFA): molecular structure and interactions with HGFA inhibitor-1 (HAI-1),” *FEBS Journal*, vol. 277, no. 10, pp. 2215-2222, 2010.

[56] S. Mukai, T. Fukushima, D. Naka, H. Tanaka, Y. Osada, and H. Kataoka, “Activation of hepatocyte growth factor activator zymogen (pro-HGFA) by human kallikrein-related peptidases,” *FEBS Journal*, vol. 275, no. 5, pp. 1003-1017, 2008.

[57] T. Shimomura et al., “Activation of the zymogen of hepatocyte growth factor activator by thrombin,” *Journal of Biological Chemistry*, vol. 268, no. 30, pp. 22927-22932, Oct. 1993.

[58] E. Bandiera et al., “Human kallikrein 5: an interesting novel biomarker in ovarian cancer patients that elicits humoral response,” *International Journal of Gynecological Cancer: Official Journal of the International Gynecological Cancer Society*, vol. 19, no. 6, pp. 1015-1021, Aug. 2009.

[59] D. Kirchhofer et al., “Tissue Expression, Protease Specificity, and Kunitz Domain Functions of Hepatocyte Growth Factor Activator Inhibitor-1B (HAI-1B), a New Splice Variant of HAI-1,” *Journal of Biological Chemistry*, vol. 278, no. 38, pp. 36341-36349, Sep. 2003.

[60] T. Shimomura, K. Denda, T. Kawaguchi, K. Matsumoto, K. Miyazawa, and N.

Kitamura, “Multiple Sites of Proteolytic Cleavage to Release Soluble Forms of Hepatocyte Growth Factor Activator Inhibitor Type 1 from a Transmembrane Form,” *Journal of Biochemistry*, vol. 126, no. 5, pp. 821-828, Nov. 1999.

[61] C. Parr and W. G. Jiang, “The role of the HGF regulatory factors in breast cancer,” in *Metastasis of Breast Cancer*, 2007, pp. 171-202.

[62] K. Miyazawa, T. Shimomura, and N. Kitamura, “Activation of Hepatocyte Growth Factor in the Injured Tissues Is Mediated by Hepatocyte Growth Factor Activator,” *Journal of Biological Chemistry*, vol. 271, no. 7, pp. 3615-3618, Feb. 1996.

[63] M. Kinosaki, K. Yamaguchi, A. Murakami, M. Ueda, T. Morinaga, and K. Higashio, “Identification of heparin-binding stretches of a naturally occurring deleted variant of hepatocyte growth factor (dHGF),” *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, vol. 1384, no. 1, pp. 93-102, Apr. 1998.

[64] W. Li, D. J. D. Johnson, C. T. Esmon, and J. A. Huntington, “Structure of the antithrombin-thrombin-heparin ternary complex reveals the antithrombotic mechanism of heparin,” *Nature Structural and Molecular Biology*, vol. 11, no. 9, pp. 857-862, 2004.

[65] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry, Fifth Edition*. W.H. Freeman, 2002.

[66] M. Lyon, J. Deakin, K. Mizuno, T. Nakamura, and J. Gallagher, “Interaction of hepatocyte growth factor with heparan sulfate. Elucidation of the major heparan sulfate structural determinants,” *Journal of Biological Chemistry*, vol. 269, no. 15, pp. 11216-11223, Apr. 1994.

[67] K. R. Catlow et al., “Interactions of Hepatocyte Growth Factor/Scatter Factor with Various Glycosaminoglycans Reveal an Important Interplay between the Presence of Iduronate and Sulfate Density,” *Journal of Biological Chemistry*, vol. 283, no. 9, pp. 5235-5248, Feb. 2008.

[68] J. A. Deakin, B. S. Blaum, J. T. Gallagher, D. Uhrín, and M. Lyon, “The Binding Properties of Minimal Oligosaccharides Reveal a Common Heparan

Sulfate/Dermatan Sulfate-binding Site in Hepatocyte Growth Factor/Scatter Factor That Can Accommodate a Wide Variety of Sulfation Patterns," *Journal of Biological Chemistry*, vol. 284, no. 10, pp. 6311 -6321, Mar. 2009.

[69] I. Capila and R. J. Linhardt, "Heparin-protein interactions," *Angewandte Chemie (International Ed. in English)*, vol. 41, no. 3, pp. 391-412, Feb. 2002.

[70] M. Lyon, J. A. Deakin, D. Lietha, E. Gherardi, and J. T. Gallagher, "The Interactions of Hepatocyte Growth Factor/Scatter Factor and Its NK1 and NK2 Variants with Glycosaminoglycans Using a Modified Gel Mobility Shift Assay: ELUCIDATION OF THE MINIMAL SIZE OF BINDING AND ACTIVATORY OLIGOSACCHARIDES," *Journal of Biological Chemistry*, vol. 279, no. 42, pp. 43560-43567, Oct. 2004.

[71] C. D. Blundell et al., "The Link Module from Ovulation- and Inflammation-associated Protein TSG-6 Changes Conformation on Hyaluronan Binding," *Journal of Biological Chemistry*, vol. 278, no. 49, pp. 49261-49270, Dec. 2003.

[72] A. Canales-Mayordomo et al., "Backbone dynamics of a biologically active human FGF-1 monomer, complexed to a hexasaccharide heparin-analogue, by 15N NMR relaxation methods," *Journal of Biomolecular NMR*, vol. 35, no. 4, pp. 225-239, 2006.

[73] T. Nakamura et al., "Molecular cloning and expression of human hepatocyte growth factor," *Nature*, vol. 342, no. 6248, pp. 440-443, Nov. 1989.

[74] T. Nakamura, K. Nawa, and A. Ichihara, "Partial purification and characterization of hepatocyte growth factor from serum of hepatectomized rats," *Biochemical and Biophysical Research Communications*, vol. 122, no. 3, pp. 1450-1459, Aug. 1984.

[75] D. Bottaro et al., "Identification of the hepatocyte growth factor receptor as the c-met proto-oncogene product," *Science*, vol. 251, no. 4995, pp. 802-804, Feb. 1991.

[76] C. S. Cooper et al., "Molecular cloning of a new transforming gene from a chemically transformed human cell line," *Nature*, vol. 311, no. 5981, pp. 29-

33, 1984.

[77] C. Birchmeier, W. Birchmeier, E. Gherardi, and G. F. Vande Woude, “Met, metastasis, motility and more,” *Nature Reviews Molecular and Cellular Biology*, vol. 4, no. 12, pp. 915-925, Dec. 2003.

[78] D. Hanahan and R. A. Weinberg, “The Hallmarks of Cancer,” *Cell*, vol. 100, no. 1, pp. 57-70, Jan. 2000.

[79] G. Maulik, A. Shrikhande, T. Kijima, P. C. Ma, P. T. Morrison, and R. Salgia, “Role of the hepatocyte growth factor receptor, c-Met, in oncogenesis and potential for therapeutic inhibition,” *Cytokine & Growth Factor Reviews*, vol. 13, no. 1, pp. 41-59, Feb. 2002.

[80] X. Liu, R. C. Newton, and P. A. Scherle, “Developing c-MET pathway inhibitors for cancer therapy: progress and challenges,” *Trends in Molecular Medicine*, vol. 16, no. 1, pp. 37-45, Jan. 2010.

[81] F. Cecchi, D. C. Rabe, and D. P. Bottaro, “Targeting the HGF/Met signalling pathway in cancer,” *European Journal of Cancer*, vol. 46, no. 7, pp. 1260-1270, May. 2010.

[82] I. J. Davis et al., “Identification of the Receptor Tyrosine Kinase c-Met and Its Ligand, Hepatocyte Growth Factor, as Therapeutic Targets in Clear Cell Sarcoma,” *Cancer Research*, vol. 70, no. 2, pp. 639-645, Jan. 2010.

[83] T. L. Underiner, T. Herbertz, and S. J. Miknyoczki, “Discovery of small molecule c-Met inhibitors: Evolution and profiles of clinical candidates,” *Anti-Cancer Agents in Medicinal Chemistry*, vol. 10, no. 1, pp. 7-27, Jan. 2010.

[84] F. Guessous et al., “An orally bioavailable c-Met kinase inhibitor potently inhibits brain tumor malignancy and growth,” *Anti-Cancer Agents in Medicinal Chemistry*, vol. 10, no. 1, pp. 28-35, Jan. 2010.

[85] T. Ueki et al., “Hepatocyte growth factor gene therapy of liver cirrhosis in rats,” *Nature Medicine*, vol. 5, no. 2, pp. 226-230, Feb. 1999.

[86] H. Jin et al., “MetMAb, the One-Armed 5D5 Anti-c-Met Antibody, Inhibits Orthotopic Pancreatic Tumor Growth and Improves Survival,” *Cancer*

Research, vol. 68, no. 11, pp. 4360-4368, Jun. 2008.

[87] M. Youles et al., "Engineering the NK1 Fragment of Hepatocyte Growth Factor/Scatter Factor as a MET Receptor Antagonist," *Journal of Molecular Biology*, vol. 377, no. 3, pp. 616-622, Mar. 2008.

[88] C. Parr, A. J. Sanders, and W. G. Jiang, "Hepatocyte growth factor activation inhibitors - therapeutic potential in cancer," *Anti-Cancer Agents in Medicinal Chemistry*, vol. 10, no. 1, pp. 47-57, Jan. 2010.

[89] http://www.vai.org/met/, "HGF/SF - Met and cancer references." [Online]. Available: http://www.vai.org/met/. [Accessed: 03-Jun-2010].

[90] R. Schwall et al., "Heparin induces dimerization and confers proliferative activity onto the hepatocyte growth factor antagonists NK1 and NK2," *Journal of Cell Biology*, vol. 133, no. 3, pp. 709-718, May. 1996.

[91] J. L. Jakubczak, W. J. Larochelle, and G. Merlino, "NK1, a Natural Splice Variant of Hepatocyte Growth Factor/Scatter Factor, Is a Partial Agonist In Vivo," *Molecular and Cellular Biology*, vol. 18, no. 3, pp. 1275-1283, Mar. 1998.

[92] T. Nakamura, K. Sakai, T. Nakamura, and K. Matsumoto, "Anti-cancer approach with NK4: Bivalent action and mechanisms," *Anti-Cancer Agents in Medicinal Chemistry*, vol. 10, no. 1, pp. 36-46, Jan. 2010.

[93] J. A. Joyce, C. Freeman, N. Meyer-Morse, C. R. Parish, and D. Hanahan, "A functional heparan sulfate mimetic implicates both heparanase and heparan sulfate in tumor angiogenesis and invasion in a mouse model of multistage cancer," *Oncogene*, vol. 24, no. 25, pp. 4037-4051, Apr. 2005.

[94] R. Lever and C. P. Page, "Novel drug development opportunities for heparin," *Nature Reviews Drug Discovery*, vol. 1, no. 2, pp. 140-148, Feb. 2002.

[95] B. Casu, I. Vlodavsky, and R. D. Sanderson, "Non-Anticoagulant Heparins and Inhibition of Cancer," *Pathophysiology, Haemostasis and Thrombosis*, vol. 36, no. 3, pp. 195-203, Jan. 2009.

[96] S. H. Kim and K. L. Kiick, "Heparin-mimetic sulfated peptides with

modulated affinities for heparin-binding peptides and growth factors," *Peptides*, vol. 28, no. 11, pp. 2125-2136, Nov. 2007.

[97] S. Vázquez-Campos, P. St. Hilaire, D. Damgaard, and M. Meldal, "GAG Mimetic Libraries: Sulphated Peptide as Heparin-like Glycosaminoglycan Mimics in Their Interaction with FGF-1," *QSAR & Combinatorial Science*, vol. 24, no. 8, pp. 923-942, 2005.

[98] V. Rouet, A. Meddahi-Pellé, H. Miao, I. Vlodavsky, J. Caruelle, and D. Barritault, "Heparin-like synthetic polymers, named RGTAs, mimic biological effects of heparin in vitro," *Journal of Biomedical Materials Research. Part A*, vol. 78, no. 4, pp. 792-7, Sep. 2006.

[99] V. Rouet et al., "A Synthetic Glycosaminoglycan Mimetic Binds Vascular Endothelial Growth Factor and Modulates Angiogenesis," *Journal of Biological Chemistry*, vol. 280, no. 38, pp. 32792-32800, Sep. 2005.

[100] I. Barbosa et al., "A synthetic glycosaminoglycan mimetic (RGTA) modifies natural glycosaminoglycan species during myogenesis," *Journal of Cellular Science*, vol. 118, no. 1, pp. 253-264, Jan. 2005.

[101] H. Vogl, R. Hoffman, D. Paper, and G. Franz, "√-and λ-carrageenan as new angiogenesis inhibitors," *Journal of Cancer Research and Clinical Oncology*, vol. 121, no. 0, p. A47, Jan. 1995.

[102] H. Chen, X. Yan, J. Lin, F. Wang, and W. Xu, "Depolymerized Products of λ-Carrageenan as a Potent Angiogenesis Inhibitor," *Journal of Agricultural and Food Chemistry*, vol. 55, no. 17, pp. 6910-6917, Aug. 2007.

[103] R. Hoffman, "Carrageenans inhibit growth-factor binding," *The Biochemical Journal*, vol. 289, pp. 331-334, Jan. 1993.

[104] R. Hoffman, W. W. Burns, and D. H. Paper, "Selective inhibition of cell proliferation and DNA synthesis by the polysulphated carbohydrate l-carrageenan," *Cancer Chemotherapy and Pharmacology*, vol. 36, no. 4, pp. 325-334, 1995.

[105] R. Hoffman and D. Sykes, "Inhibition of binding of basic fibroblast growth

factor to low and high affinity receptors by carrageenans," *Biochemical Pharmacology*, vol. 45, no. 11, pp. 2348-2351, Jun. 1993.

[106] B. Casu et al., "Short Heparin Sequences Spaced by Glycol-Split Uronate Residues Are Antagonists of Fibroblast Growth Factor 2 and Angiogenesis Inhibitors†," *Biochemistry*, vol. 41, no. 33, pp. 10519-10528, 2002.

[107] M. Presta et al., "Antiangiogenic Activity of Semisynthetic Biotechnological Heparins: Low-Molecular-Weight-Sulfated Escherichia coli K5 Polysaccharide Derivatives as Fibroblast Growth Factor Antagonists," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 25, no. 1, pp. 71-76, Jan. 2005.

[108] E. Raiber et al., "Novel heparin/heparan sulfate mimics as inhibitors of HGF/SF-induced MET activation," *Bioorganic & Medicinal Chemistry Letters*, vol. 17, no. 22, pp. 6321-6325, Nov. 2007.

[109] F. van Doormaal, H. Boller, and S. Middeldorp, "Development in anticoagulant therapy," *Critical Reviews in Oncology/Hematology*, vol. 66, no. 2, pp. 145-154, May. 2008.

[110] V. Bobek and J. Kovarík, "Antitumor and antimetastatic effect of warfarin and heparins," *Biomedecine & Pharmacotherapy*, vol. 58, no. 4, pp. 213-219, May. 2004.

[111] L. Borsig, "Antimetastatic activities of heparins and modified heparins. Experimental evidence," *Thrombosis Research*, vol. 125, no. 2, pp. S66-S71, Apr. 2010.

[112] K. Catlow et al., "Hepatocyte growth factor/scatter factor and its interaction with heparan sulphate and dermatan sulphate.," *Biochemical Society Transactions*, vol. 31, no. 2, pp. 352-353, Apr. 2003.

[113] K. Dredge et al., "The PG500 series: novel heparan sulfate mimetics as potent angiogenesis and heparanase inhibitors for cancer therapy," *Investigational New Drugs*, vol. 28, no. 3, pp. 276-283, Jun. 2010.

[114] A. H. Bartlett and P. W. Park, "Proteoglycans in host-pathogen Interactions:

Molecular Mechanisms and Therapeutic Implications," *Expert Reviews in Molecular Medicine*, vol. 12, no. 1, pp. 1-null, 2010.

[115] H. E. Bülow and O. Hobert, "The Molecular Diversity of Glycosaminoglycans Shapes Animal Development," *Annual Review of Cell and Developmental Biology*, vol. 22, no. 1, pp. 375-407, 2006.

[116] C. I. Gama et al., "Sulfation patterns of glycosaminoglycans encode molecular recognition and activity," *Nature Chemical Biology*, vol. 2, no. 9, pp. 467-473, 2006.

[117] L. Schaefer and R. Schaefer, "Proteoglycans: from structural compounds to signaling molecules," *Cell and Tissue Research*, vol. 339, no. 1, pp. 237-246, Jan. 2010.

[118] K. Matsuda et al., "Glypican-1 Is Overexpressed in Human Breast Cancer and Modulates the Mitogenic Effects of Multiple Heparin-binding Growth Factors in Breast Cancer Cells," *Cancer Research*, vol. 61, no. 14, pp. 5562-5569, Jul. 2001.

[119] A. Karihaloo, S. Kale, N. D. Rosenblum, and L. G. Cantley, "Hepatocyte Growth Factor-Mediated Renal Epithelial Branching Morphogenesis Is Regulated by Glypican-4 Expression," *Molecular and Cellular Biology*, vol. 24, no. 19, pp. 8745-8752, Oct. 2004.

[120] P. W. B. Derksen, R. M. J. Keehnen, L. M. Evers, M. H. J. van Oers, M. Spaargaren, and S. T. Pals, "Cell surface proteoglycan syndecan-1 mediates hepatocyte growth factor binding and promotes Met signaling in multiple myeloma," *Blood*, vol. 99, no. 4, pp. 1405-1410, Feb. 2002.

[121] J. D. Esko and U. Lindahl, "Molecular diversity of heparan sulfate," *Journal of Clinical Investigation*, vol. 108, no. 2, Jul. 2001.

[122] K. Hayashida, P. D. Stahl, and P. W. Park, "Syndecan-1 Ectodomain Shedding Is Regulated by the Small GTPase Rab5," *Journal of Biological Chemistry*, vol. 283, no. 51, pp. 35435 -35444, Dec. 2008.

[123] K. Sakai, T. Nakamura, K. Matsumoto, and T. Nakamura, "Angioinhibitory

Action of NK4 Involves Impaired Extracellular Assembly of Fibronectin Mediated by Perlecan-NK4 Association," *Journal of Biological Chemistry*, vol. 284, no. 33, pp. 22491-22499, 2009.

[124] A. Imberty, H. Lortat-Jacob, and S. Perez, "Structural view of glycosaminoglycan-protein interactions," *Carbohydrate Research*, vol. 342, no. 3, pp. 430-439, Feb. 2007.

[125] R. Raman, V. Sasisekharan, and R. Sasisekharan, "Structural Insights into Biological Roles of Protein-Glycosaminoglycan Interactions," *Chemistry & Biology*, vol. 12, no. 3, pp. 267-277, Mar. 2005.

[126] C. D. Nandini, T. Mikami, M. Ohta, N. Itoh, F. Akiyama-Nambu, and K. Sugahara, "Structural and Functional Characterization of Oversulfated Chondroitin Sulfate/Dermatan Sulfate Hybrid Chains from the Notochord of Hagfish," *Journal of Biological Chemistry*, vol. 279, no. 49, pp. 50799-50809, Dec. 2004.

[127] H. Mochizuki et al., "Characterization of a Heparan Sulfate 3-O-Sulfotransferase-5, an Enzyme Synthesizing a Tetrasulfated Disaccharide," *Journal of Biological Chemistry*, vol. 278, no. 29, pp. 26780 -26787, Jul. 2003.

[128] J. D. Esko and S. B. Selleck, "Order out of chaos: assembly of ligand binding sites in heparan sulfate," *Annual Review of Biochemistry*, vol. 71, pp. 435-471, 2002.

[129] M. Busse et al., "Contribution of EXT1, EXT2, and EXTL3 to Heparan Sulfate Chain Elongation," *Journal of Biological Chemistry*, vol. 282, no. 45, pp. 32802-32810, Nov. 2007.

[130] B. M. Zak, B. E. Crawford, and J. D. Esko, "Hereditary multiple exostoses and heparan sulfate polymerization," *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1573, no. 3, pp. 346-355, Dec. 2002.

[131] C. Westling and U. Lindahl, "Location of N-Unsubstituted Glucosamine Residues in Heparan Sulfate," *Journal of Biological Chemistry*, vol. 277, no. 51, pp. 49247-49255, Dec. 2002.

[132] C. Vanpouille et al., “The Heparin/Heparan Sulfate Sequence That Interacts with Cyclophilin B Contains a 3-O-Sulfated N-Unsubstituted Glucosamine Residue,” *Journal of Biological Chemistry*, vol. 282, no. 33, pp. 24416-24429, 2007.

[133] D. Shukla et al., “A Novel Role for 3-O-Sulfated Heparan Sulfate in Herpes Simplex Virus 1 Entry,” *Cell*, vol. 99, no. 1, pp. 13-22, Oct. 1999.

[134] P. Carlsson, J. Presto, D. Spillmann, U. Lindahl, and L. Kjellén, “Heparin/Heparan Sulfate Biosynthesis,” *Journal of Biological Chemistry*, vol. 283, no. 29, pp. 20008-20014, Jul. 2008.

[135] T. B. Feyerabend, J. Li, U. Lindahl, and H. Rodewald, “Heparan sulfate C5-epimerase is essential for heparin biosynthesis in mast cells,” *Nature Chemical Biology*, vol. 2, no. 4, pp. 195-196, Apr. 2006.

[136] Å. Hagner-McWhirter, J. Li, S. Oscarson, and U. Lindahl, “Irreversible Glucuronyl C5-epimerization in the Biosynthesis of Heparan Sulfate,” *Journal of Biological Chemistry*, vol. 279, no. 15, pp. 14631-14638, Apr. 2004.

[137] J. Rong, H. Habuchi, K. Kimata, U. Lindahl, and M. Kusche-Gullberg, “Substrate Specificity of the Heparan Sulfate Hexuronic Acid 2-O-Sulfotransferase†,” *Biochemistry*, vol. 40, no. 18, pp. 5548-5555, May. 2001.

[138] H. Habuchi et al., “The Occurrence of Three Isoforms of Heparan Sulfate 6-O-Sulfotransferase Having Different Specificities for Hexuronic Acid Adjacent to the TargetedN-Sulfoglucosamine,” *Journal of Biological Chemistry*, vol. 275, no. 4, pp. 2859 -2868, Jan. 2000.

[139] H. Habuchi et al., “Biosynthesis of heparan sulphate with diverse structures and functions: two alternatively spliced forms of human heparan sulphate 6-O-sulphotransferase-2 having different expression patterns and properties,” *Biochemical Journal*, vol. 371, no. 1, p. 131, 2003.

[140] L. Zhang et al., “6-O-Sulfotransferase-1 Represents a Critical Enzyme in the Anticoagulant Heparan Sulfate Biosynthetic Pathway,” *Journal of Biological Chemistry*, vol. 276, no. 45, pp. 42311-42321, Nov. 2001.

[141] B. Gorsi and S. E. Stringer, “Tinkering with heparan sulfate sulfation to steer development,” *Trends in Cell Biology*, vol. 17, no. 4, pp. 173-177, Apr. 2007.

[142] D. Xu, V. Tiwari, G. Xia, C. Clement, D. Shukla, and J. Liu, “Characterization of heparan sulphate 3-O-sulphotransferase isoform 6 and its role in assisting the entry of herpes simplex virus type 1,” *Biochemical Journal*, vol. 385, no. 2, p. 451, 2005.

[143] J. Liu et al., “Expression of Heparan Sulfate d-Glucosaminyl 3-O-Sulfotransferase Isoforms Reveals Novel Substrate Specificities,” *Journal of Biological Chemistry*, vol. 274, no. 8, pp. 5185-5192, Feb. 1999.

[144] V. Tiwari, C. D. O'Donnell, M. Oh, T. Valyi-Nagy, and D. Shukla, “A role for 3-O-sulfotransferase isoform-4 in assisting HSV-1 entry and spread,” *Biochemical and Biophysical Research Communications*, vol. 338, no. 2, pp. 930-937, Dec. 2005.

[145] J. Liu, Z. Shriver, P. Blaiklock, K. Yoshida, R. Sasisekharan, and R. D. Rosenberg, “Heparan Sulfate d-Glucosaminyl 3-O-Sulfotransferase-3A SulfatesN-Unsubstituted Glucosamine Residues,” *Journal of Biological Chemistry*, vol. 274, no. 53, pp. 38155-38162, Dec. 1999.

[146] G. Xia et al., “Heparan Sulfate 3-O-Sulfotransferase Isoform 5 Generates Both an Antithrombin-binding Site and an Entry Receptor for Herpes Simplex Virus, Type 1,” *Journal of Biological Chemistry*, vol. 277, no. 40, pp. 37912-37919, Oct. 2002.

[147] H. Mochizuki, K. Yoshida, Y. Shibata, and K. Kimata, “Tetrasulfated Disaccharide Unit in Heparan Sulfate,” *Journal of Biological Chemistry*, vol. 283, no. 45, pp. 31237-31245, Nov. 2008.

[148] G. K. Dhoot, M. K. Gustafsson, X. Ai, W. Sun, D. M. Standiford, and C. P. Emerson, “Regulation of Wnt Signaling and Embryo Patterning by an Extracellular Sulfatase,” *Science*, vol. 293, no. 5535, pp. 1663-1666, Aug. 2001.

[149] X. Ai, A. Do, O. Lozynska, M. Kusche-Gullberg, U. Lindahl, and C. P. Emerson, “QSulf1 remodels the 6-O sulfation states of cell surface heparan

sulfate proteoglycans to promote Wnt signaling," *The Journal of Cell Biology*, vol. 162, no. 2, pp. 341-351, Jul. 2003.

[150] S. Wang et al., "QSulf1, a heparan sulfate 6-O-endosulfatase, inhibits fibroblast growth factor signaling in mesoderm induction and angiogenesis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 14, pp. 4833-4838, Apr. 2004.

[151] M. Buono and M. Cosma, "Sulfatase activities towards the regulation of cell metabolism and signaling in mammals," *Cellular and Molecular Life Sciences*, vol. 67, no. 5, pp. 769-780, Mar. 2010.

[152] X. Ai, A. Do, M. Kusche-Gullberg, U. Lindahl, K. Lu, and C. P. Emerson, "Substrate Specificity and Domain Functions of Extracellular Heparan Sulfate 6-O-Endosulfatases, QSulf1 and QSulf2," *Journal of Biological Chemistry*, vol. 281, no. 8, pp. 4969-4976, Feb. 2006.

[153] E. Forsberg and L. Kjellén, "Heparan sulfate: lessons from knockout mice," *Journal of Clinical Investigation*, vol. 108, no. 2, pp. 175-180, 2001.

[154] S. Nadanaka and H. Kitagawa, "Heparan Sulphate Biosynthesis and Disease," *Journal of Biochemistry*, vol. 144, no. 1, pp. 7-14, Jul. 2008.

[155] M. Lyon and J. T. Gallagher, "Bio-specific sequences and domains in heparan sulphate and the regulation of cell growth and adhesion," *Matrix Biology*, vol. 17, no. 7, pp. 485-493, Nov. 1998.

[156] T. Toida et al., "Structural differences and the presence of unsubstituted amino groups in heparan sulphates from different tissues and species.," *Biochemical Journal*, vol. 322, no. 2, pp. 499-506, Mar. 1997.

[157] J. van den Born et al., "Novel Heparan Sulfate Structures Revealed by Monoclonal Antibodies," *Journal of Biological Chemistry*, vol. 280, no. 21, pp. 20516-20523, May. 2005.

[158] K. B. Komosińska-Vassev, K. Winsz-Szczotka, K. Kuznik-Trocha, P. Olczyk, and K. Olczyk, "Age-related changes of plasma glycosaminoglycans," *Clinical Chemistry and Laboratory Medicine: CCLM / FESCC*, vol. 46, no. 2,

pp. 219-224, 2008.

[159] K. Komosińska-Vassev, K. Winsz-Szczotka, K. Olczyk, and E. M. Koźma, “Alterations in serum glycosaminoglycan profiles in Graves' patients,” *Clinical Chemistry and Laboratory Medicine: CCLM / FESCC*, vol. 44, no. 5, pp. 582-588, 2006.

[160] P. Tatrai et al., “Quantitative and Qualitative Alterations of Heparan Sulfate in Fibrogenic Liver Diseases and Hepatocellular Cancer,” *Journal of Histochemistry and Cytochemistry*, vol. 58, no. 5, pp. 429-441, May. 2010.

[161] W. C. Lamanna, I. Kalus, M. Padva, R. J. Baldwin, C. L. Merry, and T. Dierks, “The heparanome--The enigma of encoding and decoding heparan sulfate sulfation,” *Journal of Biotechnology*, vol. 129, no. 2, pp. 290-307, Apr. 2007.

[162] U. Lindahl, M. Kusche-Gullberg, and L. Kjellen, “Regulated Diversity of Heparan Sulfate,” *Journal of Biological Chemistry*, vol. 273, no. 39, pp. 24979-24982, Sep. 1998.

[163] J. Turnbull, A. Powell, and S. Guimond, “Heparan sulfate: decoding a dynamic multifunctional cell regulator,” *Trends in Cell Biology*, vol. 11, no. 2, pp. 75-82, Feb. 2001.

[164] S. Ohtake, K. Kimata, and O. Habuchi, “Recognition of Sulfation Pattern of Chondroitin Sulfate by Uronosyl 2-O-Sulfotransferase,” *Journal of Biological Chemistry*, vol. 280, no. 47, pp. 39115-39123, Nov. 2005.

[165] X. Bao, S. Nishimura, T. Mikami, S. Yamada, N. Itoh, and K. Sugahara, “Chondroitin Sulfate/Dermatan Sulfate Hybrid Chains from Embryonic Pig Brain, Which Contain a Higher Proportion of L-Iduronic Acid than Those from Adult Pig Brain, Exhibit Neuritogenic and Growth Factor Binding Activities,” *Journal of Biological Chemistry*, vol. 279, no. 11, pp. 9765-9776, Mar. 2004.

[166] A. K. Powell, E. A. Yates, D. G. Fernig, and J. E. Turnbull, “Interactions of heparin/heparan sulfate with proteins: Appraisal of structural factors and experimental approaches,” *Glycobiology*, vol. 14, no. 4, pp. 17R-30, Apr.

2004.

[167] R. Sasisekharan, R. Raman, and V. Prabhakar, "Glycomics approach to structure-function relationships of glycosaminoglycans," *Annual Review of Biomedical Engineering*, vol. 8, pp. 181-231, 2006.

[168] J. A. Deakin and M. Lyon, "A simplified and sensitive fluorescent method for disaccharide analysis of both heparan sulfate and chondroitin/dermatan sulfates from biological samples," *Glycobiology*, vol. 18, no. 6, pp. 483-491, Jun. 2008.

[169] F. N. Lamari, M. Militsopoulou, T. N. Mitropoulou, A. Hjerpe, and N. K. Karamanos, "Analysis of glycosaminoglycan-derived disaccharides in biologic samples by capillary electrophoresis and protocol for sequencing glycosaminoglycans," *Biomedical Chromatography*, vol. 16, no. 2, pp. 95-102, 2002.

[170] A. K. Korir, J. F. K. Limtiaco, S. M. Gutierrez, and C. K. Larive, "Ultraperformance Ion-Pair Liquid Chromatography Coupled to Electrospray Time-of-Flight Mass Spectrometry for Compositional Profiling and Quantification of Heparin and Heparan Sulfate," *Analytical Chemistry*, vol. 80, no. 4, pp. 1297-1306, Feb. 2008.

[171] O. M. Saad, H. Ebel, K. Uchimura, S. D. Rosen, C. R. Bertozzi, and J. A. Leary, "Compositional profiling of heparin/heparan sulfate using mass spectrometry: assay for specificity of a novel extracellular human endosulfatase," *Glycobiology*, vol. 15, no. 8, pp. 818-826, 2005.

[172] B. Mulloy and M. J. Forster, "Conformation and dynamics of heparin and heparan sulfate," *Glycobiology*, vol. 10, no. 11, pp. 1147-1156, Nov. 2000.

[173] B. Mulloy, M. J. Forster, C. Jones, and D. B. Davies, "N.m.r. and molecular-modelling studies of the solution conformation of heparin.," *Biochemical Journal*, vol. 293, no. 3, pp. 849-858, Aug. 1993.

[174] D. R. Coombe and W. C. Kett, "Heparan sulfate-protein interactions: therapeutic potential through structure-function insights," *Cellular and Molecular Life Sciences (CMLS)*, vol. 62, no. 4, pp. 410-424, Feb. 2005.

[175] J. L. Stevenson, A. Varki, and L. Borsig, "Heparin attenuates metastasis mainly due to inhibition of P- and L-selectin, but non-anticoagulant heparins can have additional effects," *Thrombosis Research*, vol. 120, no. 2, pp. S107-S111, 2007.

[176] L. Wang, J. R. Brown, A. Varki, and J. D. Esko, "Heparin's anti-inflammatory effects require glucosamine 6-O-sulfation and are mediated by blockade of L- and P-selectins," *Journal of Clinical Investigation*, vol. 110, no. 1, pp. 127-136, 2002.

[177] H. Lortat-Jacob, A. Grosdidier, and A. Imberty, "Structural diversity of heparan sulfate binding domains in chemokines," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 3, pp. 1229-1234, Feb. 2002.

[178] C. R. Parish, "The role of heparan sulphate in inflammation," *Nat Rev Immunol*, vol. 6, no. 9, pp. 633-643, 2006.

[179] Y. Nadir and B. Brenner, "Heparanase coagulation and cancer progression," *Best Practice & Research Clinical Haematology*, vol. 22, no. 1, pp. 85-92, Mar. 2009.

[180] I. Vlodavsky et al., "The impact of heparanese and heparin on cancer metastasis and angiogenesis," *Pathophysiology of Haemostasis and Thrombosis*, vol. 35, no. 1, pp. 116-127, 2006.

[181] I. Vlodavsky et al., "Heparanase, heparin and the coagulation system in cancer progression," *Thrombosis Research*, vol. 120, no. 2, pp. S112-S120, 2007.

[182] O. Ben-Zaken, S. Gingis-Velitski, I. Vlodavsky, and N. Ilan, "Heparanase induces Akt phosphorylation via a lipid raft receptor," *Biochemical and biophysical research communications*, vol. 361, no. 4, pp. 829-834, Oct. 2007.

[183] J. Li and I. Vlodavsky, "Heparin, heparan sulfate and heparanase in inflammatory reactions," *Thrombosis and Haemostasis*, vol. 102, no. 5, pp. 823-828, Nov. 2009.

[184] S. Kolset and H. Tveit, "Serglycin – Structure and biology," *Cellular and*

Molecular Life Sciences, vol. 65, no. 7, pp. 1073-1085, Apr. 2008.

[185] R. Lever, A. Smailbegovic, and C. P. Page, "Locally available heparin modulates inflammatory cell recruitment in a manner independent of anticoagulant activity," *European Journal of Pharmacology*, vol. 630, no. 1, pp. 137-144, Mar. 2010.

[186] T. Ahmed, J. Garrigo, and I. Danta, "Preventing Bronchoconstriction in Exercise-Induced Asthma with Inhaled Heparin," *New England Journal of Medicine*, vol. 329, no. 2, pp. 90-95, Jul. 1993.

[187] L. Borsig, "Antimetastatic Activities of Modified Heparins: Selectin Inhibition by Heparin Attenuates Metastasis," *Seminars in Thrombosis and Hemostasis*, vol. 33, no. 5, pp. 540-546, 2007.

[188] R. E. Hileman, J. R. Fromm, J. M. Weiler, and R. J. Linhardt, "Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins," *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, vol. 20, no. 2, pp. 156-167, Feb. 1998.

[189] E. Seyrek and P. Dubin, "Glycosaminoglycans as polyelectrolytes," *Advances in Colloid and Interface Science*, vol. 158, no. 1, pp. 119-129, Jul. 2010.

[190] J. Gummel, F. Cousin, and F. Boué, "Counterions Release from Electrostatic Complexes of Polyelectrolytes and Proteins of Opposite Charge: A Direct Measurement," *Journal of the American Chemical Society*, vol. 129, no. 18, pp. 5806-5807, May. 2007.

[191] K. A. Schug and W. Lindner, "Noncovalent Binding between Guanidinium and Anionic Groups: Focus on Biological- and Synthetic-Based Arginine/Guanidinium Interactions with Phosph[on]ate and Sulf[on]ate Residues," *Chemical Reviews*, vol. 105, no. 1, pp. 67-114, Jan. 2005.

[192] D. Fang, P. Fabian, Z. Szekely, X. Fu, T. Tang, and I. G. Csizmadia, "Structure and stability of ammonium-sulfate and guanidium-sulfate complex," *Journal of Molecular Structure: THEOCHEM*, vol. 430, pp. 161-170, Apr. 1998.

[193] J. R. Fromm, R. E. Hileman, E. E. O. Caldwell, J. M. Weiler, and R. J. Linhardt, "Differences in the Interaction of Heparin with Arginine and Lysine and the Importance of these Basic Amino Acids in the Binding of Heparin to Acidic Fibroblast Growth Factor," *Archives of Biochemistry and Biophysics*, vol. 323, no. 2, pp. 279-287, Nov. 1995.

[194] R. Raman, G. Venkataraman, S. Ernst, V. Sasisekharan, and R. Sasisekharan, "Structural specificity of heparin binding in the fibroblast growth factor family of proteins," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2357-2362, Mar. 2003.

[195] T. Rudd et al., "Site-specific interactions of copper(II) ions with heparin revealed with complementary (SRCD, NMR, FTIR and EPR) spectroscopic techniques," *Carbohydrate Research*, vol. 343, no. 12, pp. 2184-2193, Aug. 2008.

[196] D. Grant, W. F. Long, C. F. Moffat, and F. B. Williamson, "A study of Ca(2+)-heparin complex-formation by polarimetry.," *Biochemical Journal*, vol. 282, no. 2, pp. 601-604, Mar. 1992.

[197] T. R. Rudd et al., "Influence of substitution pattern and cation binding on conformation and activity in heparin derivatives," *Glycobiology*, vol. 17, no. 9, pp. 983-993, Sep. 2007.

[198] R. Hjelm and S. Schedin-Weiss, "High Affinity Interaction between a Synthetic, Highly Negatively Charged Pentasaccharide and α - or β - Antithrombin Is Predominantly Due to Nonionic Interactions†," *Biochemistry*, vol. 46, no. 11, pp. 3378-3384, Mar. 2007.

[199] J. I. Weitz, "New Anticoagulants for Treatment of Venous Thromboembolism," *Circulation*, vol. 110, no. 9, pp. I-19-26, Aug. 2004.

[200] M. Yanagishita, K. Podyma-Inoue, and M. Yokoyama, "Extraction and separation of proteoglycans," *Glycoconjugate Journal*, vol. 26, no. 8, pp. 953-959, Nov. 2009.

[201] S. Yamada et al., "Determination of the Glycosaminoglycan-Protein Linkage Region Oligosaccharide Structures of Proteoglycans from Drosophila

melanogaster and Caenorhabditis elegans," *Journal of Biological Chemistry*, vol. 277, no. 35, pp. 31877-31886, 2002.

[202] J. L. D. Paz and P. H. Seeberger, "Deciphering the glycosaminoglycan code with the help of microarrays," *Molecular BioSystems*, vol. 4, no. 7, pp. 707-711.

[203] T. Toida, K. Sato, N. Sakamoto, S. Sakai, S. Hosoyama, and R. J. Linhardt, "Solvolytic depolymerization of chondroitin and dermatan sulfates," *Carbohydrate Research*, vol. 344, no. 7, pp. 888-893, May. 2009.

[204] P. L. Steyn, P. Segers, M. Vancanneyt, P. Sandra, K. Kersters, and J. J. Joubert, "Classification of heparinolytic bacteria into a new genus, Pedobacter, comprising four species: Pedobacter heparinus comb. nov., Pedobacter piscium comb. nov., Pedobacter africanus sp. nov. and Pedobacter saltans sp. nov. Proposal of the family Sphingobacteriaceae fam. nov.," *International Journal of Systematic Bacteriology*, vol. 48, no. 1, pp. 165-177, Jan. 1998.

[205] M. A. Kertesz, "Riding the sulfur cycle metabolism of sulfonates and sulfate esters in Gram-negative bacteria," *FEMS Microbiology Reviews*, vol. 24, no. 2, pp. 135-175, 2000.

[206] D. L. Lohse and R. J. Linhardt, "Purification and characterization of heparin lyases from Flavobacterium heparinum.," *Journal of Biological Chemistry*, vol. 267, no. 34, pp. 24347-24355, Dec. 1992.

[207] D. Shaya et al., "Catalytic mechanism of heparinase II investigated by site-directed mutagenesis and the crystal structure with its substrate," *Journal of Biological Chemistry*, 2010.

[208] R. J. Linhardt, K. G. Rice, Y. S. Kim, D. L. Lohse, H. M. Wang, and D. Loganathan, "Mapping and quantification of the major oligosaccharide components of heparin.," *Biochemical Journal*, vol. 254, no. 3, Sep. 1988.

[209] Z. Shriver, D. Liu, Y. Hu, and R. Sasisekharan, "Biochemical Investigations and Mapping of the Calcium-binding Sites of Heparinase I from Flavobacterium heparinum," *Journal of Biological Chemistry*, vol. 274, no. 7, pp. 4082-4088, Feb. 1999.

[210] U. R. Desai, H. M. Wang, and R. J. Linhardt, "Substrate Specificity of the Heparin Lyases from Flavobacterium heparinum," *Archives of Biochemistry and Biophysics*, vol. 306, no. 2, pp. 461-468, Nov. 1993.

[211] V. Prabhakar, R. Raman, I. Capila, C. Bosques, K. Pojasek, and R. Sasisekharan, "Biochemical characterization of the chondroitinase ABC I active site," *Biochemical Journal*, vol. 390, no. 2, p. 395, 2005.

[212] J. Féthière, B. Eggimann, and M. Cygler, "Crystal structure of chondroitin AC lyase, a representative of a family of glycosaminoglycan degrading enzymes," *Journal of Molecular Biology*, vol. 288, no. 4, pp. 635-647, May. 1999.

[213] W. Huang et al., "Crystal structure of chondroitinase B from Flavobacterium heparinum and its complex with a disaccharide product at 1.7 Å resolution," *Journal of Molecular Biology*, vol. 294, no. 5, pp. 1257-1269, Dec. 1999.

[214] Sigma-Aldrich, "BioFiles - Complex Carbohydrate Analysis: Enzymes, Kits and Reagents," vol. 2, no. 3, 2007.

[215] Iduron, "Iduron technical bulletin - Bacterial Heparinase Enzymes Heparin and Heparan Sulfate Lyases from Flavobacterium heparinum."

[216] B. Tissot et al., "Towards GAG glycomics: Analysis of highly sulfated heparins by MALDI-TOF mass spectrometry," *Glycobiology*, vol. 17, no. 9, pp. 972-982, Sep. 2007.

[217] W. F. Vranken et al., "The CCPN data model for NMR spectroscopy: Development of a software pipeline," *Proteins: Structure, Function, and Bioinformatics*, vol. 59, no. 4, pp. 687-696, 2005.

[218] Y. Dai, R. M. Whittal, C. A. Bridges, Y. Isogai, O. Hindsgaul, and L. Li, "Matrix-assisted laser desorption ionization mass spectrometry for the analysis of monosulfated oligosaccharides," *Carbohydrate Research*, vol. 304, no. 1, pp. 1-9, Oct. 1997.

[219] D. I. Papac, A. Wong, and A. J. S. Jones, "Analysis of Acidic Oligosaccharides and Glycopeptides by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry," *Analytical*

Chemistry, vol. 68, no. 18, pp. 3215-3223, Jan. 1996.

[220] P. Juhasz and K. Biemann, "Mass spectrometric molecular-weight determination of highly acidic compounds of biological significance via their complexes with basic polypeptides.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 10, pp. 4333–4337, May. 1994.

[221] P. Juhasz and K. Biemann, "Utility of non-covalent complexes in the matrix-assisted laser desorption ionization mass spectrometry of heparin-derived oligosaccharides," *Carbohydrate Research*, vol. 270, no. 2, pp. 131-147, Apr. 1995.

[222] Invitrogen, *Pichia Expression Kit, Version M, 011102, 25-0043*. .

[223] M. J. Czar, J. C. Anderson, J. S. Bader, and J. Peccoud, "Gene synthesis demystified," *Trends in Biotechnology*, vol. 27, no. 2, pp. 63-72, Feb. 2009.

[224] S. E. Harding and J. C. Horton, "Computer-aided interpretation of analytical sedimentation data for proteins," in *Analytical ultracentrifugation in biochemistry and polymer science*, Royal Society of Chemistry, 1992, pp. 90-125.

[225] H. Hinz, "Specific Volumes of Biological Macromolecules and Some other Molecules of Biological Interest," in *Thermodynamic data for biochemistry and biotechnology*, Springer-Verlag, 1986, pp. 45-128.

[226] G. H. Barlow, N. D. Sanderson, and P. D. McNeill, "Macromolecular properties and biological activity of heparin," *Archives of Biochemistry and Biophysics*, vol. 94, no. 3, pp. 518-525, Sep. 1961.

[227] S. E. Lasker and S. S. Stivala, "Physicochemical studies of fractionated bovine heparin : I. Some dilute solution properties," *Archives of Biochemistry and Biophysics*, vol. 115, no. 2, pp. 360-372, Aug. 1966.

[228] P. Schuck, "Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation.," *Biophysical Journal*, vol. 75, no. 3, pp. 1503-1512, Sep. 1998.

[229] O. Lamm, *Die Differentialgleichung der Ultrazentrifugierung*. Almqvist & Wiksell, 1929.

[230] P. Schuck, "Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling," *Biophysical Journal*, vol. 78, no. 3, pp. 1606-1619, 2000.

[231] P. Schuck, "On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation," *Analytical Biochemistry*, vol. 320, no. 1, pp. 104-124, Sep. 2003.

[232] P. Brown and P. Schuck, "Macromolecular Size-and-Shape Distributions by Sedimentation Velocity Analytical Ultracentrifugation," *Biophysical Journal*, vol. 90, no. 12, pp. 4651-4661, 2006.

[233] J. García De La Torre, M. L. Huertas, and B. Carrasco, "Calculation of hydrodynamic properties of globular proteins from their atomic-level structure.," *Biophysical Journal*, vol. 78, no. 2, pp. 719-730, Feb. 2000.

[234] T. L. Hwang and A. J. Shaka, "Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients," *Journal of Magnetic Resonance, Series A*, vol. 112, no. 2, pp. 275-279, Feb. 1995.

[235] N. A. Farrow et al., "Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation," *Biochemistry*, vol. 33, no. 19, pp. 5984-6003, 1994.

[236] N. C. Rockwell and R. S. Fuller, "Interplay between S1 and S4 Subsites in Kex2 Protease: Kex2 Exhibits Dual Specificity for the P4 Side Chain†," *Biochemistry*, vol. 37, no. 10, pp. 3386-3391, Mar. 1998.

[237] M. Jahic, A. Veide, T. Charoenrat, T. Teeri, and S. Enfors, "Process Technology for Production and Recovery of Heterologous Proteins with Pichia pastoris," *Biotechnology Progress*, vol. 22, no. 6, pp. 1465-1473, Dec. 2006.

[238] M. Jahic, F. Wallberg, M. Bollok, P. Garcia, and S. Enfors, "Temperature limited fed-batch technique for control of proteolysis in Pichia pastoris bioreactor cultures," *Microbial Cell Factories*, vol. 2, no. 1, p. 6, 2003.

[239] A. Murasugi, Y. Asami, and Y. Mera-Kikuchi, “Production of Recombinant Human Bile Salt-Stimulated Lipase in Pichia pastoris,” *Protein Expression and Purification*, vol. 23, no. 2, pp. 282-288, Nov. 2001.

[240] J. L. Cereghino and J. M. Cregg, “Heterologous protein expression in the methylotrophic yeast Pichia pastoris,” *FEMS Microbiology Reviews*, vol. 24, no. 1, pp. 45-66, Jan. 2000.

[241] R. A. Brierley, “Secretion of Recombinant Human Insulin-Like Growth Factor I (IGF-I),” in *Pichia Protocols*, 1998, pp. 149-177.

[242] M. Trucksis, T. Conn, A. Fasano, and J. Kaper, “Production of Vibrio cholerae accessory cholera enterotoxin (Ace) in the yeast Pichia pastoris,” *Infection and Immunology*, vol. 65, no. 12, pp. 4984-4988, Dec. 1997.

[243] R. L. Tuinstra, F. C. Peterson, S. Kutlesa, E. S. Elgin, M. A. Kron, and B. F. Volkman, “Interconversion between two unrelated protein folds in the lymphotactin native state,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 13, pp. 5057-5062, Apr. 2008.

[244] A. Sinz, “Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions,” *Mass Spectrometry Reviews*, vol. 25, no. 4, pp. 663-682, 2006.

[245] T. N. Huckerby, I. A. Nieduszynski, M. Giannopoulos, S. D. Weeks, I. H. Sadler, and R. M. Lauder, “Characterization of oligosaccharides from the chondroitin/dermatan sulfates,” *FEBS Journal*, vol. 272, no. 24, pp. 6276-6286, 2005.

[246] K. Sugahara et al., “Chondroitinase ABC-resistant sulfated trisaccharides isolated from digests of chondroitin/dermatan sulfate chains,” *Carbohydrate Research*, vol. 255, pp. 165-182, Mar. 1994.

[247] L. Jin, “Investigating Glycosaminoglycan Conformations by Ion Mobiity Mass Spectrometry and Nuclear Magnetic Resonance.”

[248] Z. M. Merchant, Y. S. Kim, K. G. Rice, and R. J. Linhardt, “Structure of heparin-derived tetrasaccharides,” *The Biochemical Journal*, vol. 229, no. 2,

pp. 369-377, Jul. 1985.

[249] S. Murugesan, J. Xie, and R. J. Linhardt, "Immobilization of heparin: approaches and applications," *Current Topics in Medicinal Chemistry*, vol. 8, no. 2, pp. 80-100, 2008.

[250] Thermo Scientific, "Thermo Scientific Avidin-Biotin Technical Handbook," 07-May-2010. [Online]. Available: http://www.piercenet.com/files/1601675_AvBi_HB_INTL.pdf. [Accessed: 07-May-2010].

[251] C. H. Grün et al., "One-step biotinylation procedure for carbohydrates to study carbohydrate-protein interactions," *Analytical Biochemistry*, vol. 354, no. 1, pp. 54-63, Jul. 2006.

[252] J. C. Gildersleeve, O. Oyelaran, J. T. Simpson, and B. Allred, "Improved Procedure for Direct Coupling of Carbohydrates to Proteins via Reductive Amination," *Bioconjugate Chemistry*, vol. 19, no. 7, pp. 1485-1490, Jul. 2008.

[253] B. L. Ridley, M. D. Spiro, J. Glushka, P. Albersheim, A. Darvill, and D. Mohnen, "A Method for Biotin Labeling of Biologically Active Oligogalacturonides Using a Chemically Stable Hydrazide Linkage," *Analytical Biochemistry*, vol. 249, no. 1, pp. 10-19, Jun. 1997.

[254] J. Hsu, S. J. Chang, and A. H. Franz, "MALDI-TOF and ESI-MS Analysis of Oligosaccharides Labeled with a New Multifunctional Oligosaccharide Tag," *Journal of the American Society for Mass Spectrometry*, vol. 17, no. 2, pp. 194-204, Feb. 2006.

[255] M. D. Spiro et al., "Biological Activity of Reducing-End-Derivatized Oligogalacturonides in Tobacco Tissue Cultures," *Plant Physiology*, vol. 116, no. 4, pp. 1289–1298, Apr. 1998.

[256] Y. Shinohara et al., "Bifunctional Labeling Reagent for Oligosaccharides To Incorporate Both Chromophore and Biotin Groups," *Analytical Chemistry*, vol. 68, no. 15, pp. 2573-2579, Jan. 1996.

[257] C. Leteux, R. Childs, W. Chai, M. Stoll, H. Kogelberg, and T. Feizi, "Biotinyl-

l-3-(2-naphthyl)-alanine hydrazide derivatives of N-glycans: versatile solid-phase probes for carbohydrate-recognition studies," *Glycobiology*, vol. 8, no. 3, pp. 227-236, Mar. 1998.

[258] Y. Shinohara et al., "Use of a Biosensor Based on Surface Plasmon Resonance and Biotinyl Glycans for Analysis of Sugar Binding Specificities of Lectins," *Journal of Biochemistry*, vol. 117, no. 5, pp. 1076-1082, May. 1995.

[259] P. Kapková, "Mass spectrometric analysis of carbohydrates labeled with a biotinylated tag," *Rapid Communications in Mass Spectrometry*, vol. 23, no. 17, pp. 2775-2784, 2009.

[260] B. Kuberan, N. Gunay, J. Dordick, and R. Linhardt, "Preparation and isolation of neoglycoconjugates using biotin-streptavidin complexes," *Glycoconjugate Journal*, vol. 16, no. 6, pp. 271-281, Jun. 1999.

[261] D. S. Dalpathado, H. Jiang, M. A. Kater, and H. Desaire, "Reductive amination of carbohydrates using NaBH(OAc)3," *Analytical and Bioanalytical Chemistry*, vol. 381, no. 6, pp. 1130-1137, Mar. 2005.

[262] Y. Hatanaka, U. Kempin, and P. Jong-Jip, "One-Step Synthesis of Biotinyl Photoprobes from Unprotected Carbohydrates," *The Journal of Organic Chemistry*, vol. 65, no. 18, pp. 5639-5643, 2000.

[263] L. Kinsella, H. Chen, J. A. Smith, P. S. Rudland, and D. G. Fernig, "Interactions of putative heparin-binding domains of basic fibroblast growth factor and its receptor, FGFR-1, with heparin using synthetic peptides," *Glycoconjugate Journal*, vol. 15, no. 4, pp. 419-422, Apr. 1998.

[264] A. K. Powell, D. G. Fernig, and J. E. Turnbull, "Fibroblast Growth Factor Receptors 1 and 2 Interact Differently with Heparin/Heparan Sulfate," *Journal of Biological Chemistry*, vol. 277, no. 32, pp. 28554-28563, 2002.

[265] H. Sakata et al., "Heparin Binding and Oligomerization of Hepatocyte Growth Factor/Scatter Factor Isoforms. HEPARAN SULFATE GLYCOSAMINOGLYCAN REQUIREMENT FOR Met BINDING AND SIGNALING," *Journal of Biological Chemistry*, vol. 272, no. 14, pp. 9457-9463, Apr. 1997.

[266] S. Ashikari, H. Habuchi, and K. Kimata, “Characterization of Heparan Sulfate Oligosaccharides That Bind to Hepatocyte Growth Factor,” *J. Biol. Chem.*, vol. 270, no. 49, pp. 29586-29593, Dec. 1995.

[267] S. Yamauchi et al., “Phosphorylation of Neuroglycan C, a Brain-specific Transmembrane Chondroitin Sulfate Proteoglycan, and Its Localization in the Lipid Rafts,” *Journal of Biological Chemistry*, vol. 277, no. 23, pp. 20583-20590, Jun. 2002.

[268] G. T. Hermanson, *Bioconjugate techniques*. Academic Press, 2008.

[269] N. S. Gandhi and R. L. Mancera, “The Structure of Glycosaminoglycans and their Interactions with Proteins,” *Chemical Biology & Drug Design*, vol. 72, no. 6, pp. 455-482, 2008.

[270] N. J. Harmer et al., “Multimers of the fibroblast growth factor (FGF)–FGF receptor–saccharide complex are formed on long oligomers of heparin,” *Biochemical Journal*, vol. 393, no. 3, pp. 741-748, Feb. 2006.

[271] G. Waksman and A. B. Herr, “New insights into heparin-induced FGF oligomerization,” *Nat Struct Mol Biol*, vol. 5, no. 7, pp. 527-530, Jul. 1998.

[272] M. K. Pangburn, N. Rawal, C. Cortes, M. N. Alam, V. P. Ferreira, and M. A. L. Atkinson, “Polyanion-Induced Self-Association of Complement Factor H,” *J Immunol*, vol. 182, no. 2, pp. 1061-1068, Jan. 2009.

[273] D. Svergun, C. Barberato, and M. H. J. Koch, “CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates,” *Journal of Applied Crystallography*, vol. 28, no. 6, pp. 768-773, 1995.

[274] P. H. Brown and P. Schuck, “Macromolecular Size-and-Shape Distributions by Sedimentation Velocity Analytical Ultracentrifugation,” *Biophysical Journal*, vol. 90, no. 12, pp. 4651-4661, Jun. 2006.

[275] D. J. Scott, S. E. Harding, A. J. Rowe, and R. S. O. C. (. Britain), *Analytical ultracentrifugation: techniques and methods*. Royal Society of Chemistry, 2005.

[276] H. Rahmoune, P. S. Rudland, J. T. Gallagher, and D. G. Fernig, “Hepatocyte Growth Factor/Scatter Factor Has Distinct Classes of Binding Site in Heparan Sulfate from Mammary Cells†,” *Biochemistry*, vol. 37, no. 17, pp. 6003-6008, Apr. 1998.

[277] R. R. Vives, R. Sadir, A. Imberty, A. Rencurosi, and H. Lortat-Jacob, “A Kinetics and Modeling Study of RANTES(9−68) Binding to Heparin Reveals a Mechanism of Cooperative Oligomerization,” *Biochemistry*, vol. 41, no. 50, pp. 14779-14789, Dec. 2002.

[278] R. I. W. Osmond, W. C. Kett, S. E. Skett, and D. R. Coombe, “Protein-heparin interactions measured by BIAcore 2000 are affected by the method of heparin immobilization,” *Analytical Biochemistry*, vol. 310, no. 2, pp. 199-207, Nov. 2002.

[279] R. L. Rich et al., “A global benchmark study using affinity-based biosensors,” *Analytical Biochemistry*, vol. 386, no. 2, pp. 194-216, Mar. 2009.

[280] M. M. Pierce, C. S. Raman, and B. T. Nall, “Isothermal Titration Calorimetry of Protein-Protein Interactions,” *Methods*, vol. 19, no. 2, pp. 213-221, Oct. 1999.

[281] I. Ahl, B. Jonsson, and L. A. E. Tibell, “Thermodynamic Characterization of the Interaction between the C-Terminal Domain of Extracellular Superoxide Dismutase and Heparin by Isothermal Titration Calorimetry,” *Biochemistry*, vol. 48, no. 41, pp. 9932-9940, Oct. 2009.

[282] G. A. Holdgate, “Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics,” *BioTechniques*, vol. 31, no. 1, pp. 164-166, 168, 170 passim, Jul. 2001.

[283] M. Lo et al., “Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery,” *Analytical Biochemistry*, vol. 332, no. 1, pp. 153-159, Sep. 2004.

[284] P. Cimmperman et al., “A Quantitative Model of Thermal Stabilization and Destabilization of Proteins by Ligands,” *Biophysical Journal*, vol. 95, no. 7, pp. 3222-3231, Oct. 2008.

[285] E. R. P. Zuiderweg, “Mapping Protein−Protein Interactions in Solution by NMR Spectroscopy†,” *Biochemistry*, vol. 41, no. 1, pp. 1-7, Jan. 2002.

[286] Sachchidanand et al., “Mapping the Heparin-binding Site on the13–14F3 Fragment of Fibronectin,” *Journal of Biological Chemistry*, vol. 277, no. 52, pp. 50629 -50635, Dec. 2002.

[287] E. Raulo et al., “The Two Thrombospondin Type I Repeat Domains of the Heparin-binding Growth-associated Molecule Bind to Heparin/Heparan Sulfate and Regulate Neurite Extension and Plasticity in Hippocampal Neurons,” *Journal of Biological Chemistry*, vol. 280, no. 50, pp. 41576 - 41583, Dec. 2005.

[288] S. Kuo-Wei Hung et al., “Solution Structure of the Ligand Binding Domain of the Fibroblast Growth Factor Receptor: Role of Heparin in the Activation of the Receptor†,‡,” 09-Nov-2005. [Online]. Available: http://pubs.acs.org/doi/abs/10.1021/bi051030n. [Accessed: 10-Jan-2009].

[289] C. Q. Schmidt et al., “A New Map of Glycosaminoglycan and C3b Binding Sites on Factor H,” *Journal of Immunology*, vol. 181, no. 4, pp. 2610-2619, Aug. 2008.

[290] N. S. Gunay, K. Tadano-Aritomi, T. Toida, I. Ishizuka, and R. J. Linhardt, “Evaluation of Counterions for Electrospray Ionization Mass Spectral Analysis of A Highly Sulfated Carbohydrate, Sucrose Octasulfate,” *Analytical Chemistry*, vol. 75, no. 13, pp. 3226-3231, Jul. 2003.

[291] B. E. Prosser et al., “Structural basis for complement factor H–linked age-related macular degeneration,” *The Journal of Experimental Medicine*, vol. 204, no. 10, pp. 2277-2283, Oct. 2007.

[292] C. A. Lepre, J. M. Moore, and J. W. Peng, “Theory and Applications of NMR-Based Screening in Pharmaceutical Research,” *Chemical Reviews*, vol. 104, no. 8, pp. 3641-3676, 2004.

[293] J. Krishnamoorthy, V. C. K. Yu, and Y. Mok, “Auto-FACE: An NMR Based Binding Site Mapping Program for Fast Chemical Exchange Protein-Ligand Systems,” vol. 5, no. 2.

[294] J. Clarkson and I. D. Campbell, “Studies of protein-ligand interactions by NMR,” *Biochemical Society Transactions*, vol. 31, no. 5, pp. 1006-1009, Oct. 2003.

[295] L. Fielding, S. Rutherford, and D. Fletcher, “Determination of protein-ligand binding affinity by NMR: observations from serum albumin model systems,” *Magnetic Resonance in Chemistry: MRC*, vol. 43, no. 6, pp. 463-470, Jun. 2005.

[296] H. O. Villar, J. Yan, and M. R. Hansen, “Using NMR for ligand discovery and optimization,” *Current Opinion in Chemical Biology*, vol. 8, no. 4, pp. 387-391, Aug. 2004.

[297] S. B. Shuker, P. J. Hajduk, R. P. Meadows, and S. W. Fesik, “Discovering High-Affinity Ligands for Proteins: SAR by NMR,” *Science*, vol. 274, no. 5292, pp. 1531-1534, Nov. 1996.

[298] B. S. Blaum et al., “Lysine and arginine side chains in glycosaminoglycan-protein complexes investigated by NMR, cross-linking, and mass spectrometry: a case study of the factor H-heparin interaction,” *Journal of the American Chemical Society*, vol. 132, no. 18, pp. 6374-6381, May. 2010.

[299] E. T. Mack, R. Perez-Castillejos, Z. Suo, and G. M. Whitesides, “Exact Analysis of Ligand-Induced Dimerization of Monomeric Receptors,” *Analytical Chemistry*, vol. 80, no. 14, pp. 5550-5555, Jul. 2008.

[300] G. C. K. Roberts, *NMR of macromolecules: a practical approach*. Oxford University Press, 1993.

[301] N. U. Jain, A. Venot, K. Umemoto, H. Leffler, and J. H. Prestegard, “Distance mapping of protein-binding sites using spin-labeled oligosaccharide ligands,” *Protein Science: A Publication of the Protein Society*, vol. 10, no. 11, pp. 2393-2400, Nov. 2001.

[302] P. E. Johnson, E. Brun, L. F. MacKenzie, S. G. Withers, and L. P. McIntosh, “The cellulose-binding domains from Cellulomonas fimi [beta]-1,4-glucanase CenC bind nitroxide spin-labeled cellooligosaccharides in multiple orientations,” *Journal of Molecular Biology*, vol. 287, no. 3, pp. 609-625, Apr.

1999.

[303] M. Forster and B. Mulloy, “Computational approaches to the identification of heparin-binding sites on the surfaces of proteins.,” *Biochemical Society Transactions*, vol. 34, no. 3, pp. 431-434, Jun. 2006.

[304] E. Gemma, O. Meyer, D. Uhrin, and A. N. Hulme, “Enabling methodology for the end functionalisation of glycosaminoglycan oligosaccharides,” *Molecular BioSystems*, vol. 4, no. 6, pp. 481-495.

[305] A. V. Gudmundsdottir, C. E. Paul, and M. Nitz, “Stability studies of hydrazide and hydroxylamine-based glycoconjugates in aqueous solution,” *Carbohydrate Research*, vol. 344, no. 3, pp. 278-284, Feb. 2009.

[306] A. Ido and H. Tsubouchi, “Translational research to identify clinical applications of hepatocyte growth factor,” *Hepatology Research*, vol. 39, no. 8, pp. 739-747, 2009.

[307] S. Yasuda et al., “Single Low-Dose Administration of Human Recombinant Hepatocyte Growth Factor Attenuates Intimal Hyperplasia in a Balloon-Injured Rabbit Iliac Artery Model,” *Circulation*, vol. 101, no. 21, pp. 2546-2549, May. 2000.

[308] Y. Cui et al., “Recombinant human hepatocyte growth factor for liver failure,” *Contemporary Clinical Trials*, vol. 29, no. 5, pp. 696-704, 2008.

[309] S. Yoshida et al., “Recombinant Hepatocyte Growth Factor Accelerates Cutaneous Wound Healing in a Diabetic Mouse Model,” *Growth Factors*, vol. 22, no. 2, p. 111, 2004.

[310] A. Ido et al., “Pharmacokinetic study of recombinant human hepatocyte growth factor administered in a bolus intravenously or via portal vein,” *Hepatology Research*, vol. 30, no. 3, pp. 175-181, Nov. 2004.